Nonlinear Control

This chapter will focus on the topic of nonlinear control for nonlinear system behaves naturally. Linear system is only an approximation to our life. In this chapter,modeling and solution of nonlinear control system will be provided to design the gain of nonlinear control. Furthermore, nonlinear optimal control and nonlinear observer are also explained to help readers analyze and design a nonlinear system. Besides these contents, the linearization methods of nonlinear control system are also provided for readers to approximately analyze a nonlinear system.

Objectives

When you have finished this chapter, you should be able to:

- Learn some basic concepts and properties on nonlinear control.
- Know how to convert a nonlinear control system into an linear system.
- Understand the controllability and observability of nonlinear control
- Grasp the idea of nonlinear gain control and nonlinear optimal control.
- Know a nonlinear observer.

12.1 Introduction

Usually control system is always described by input variables u or r and output variable y. In this sense, a control system is considered as a mapping from the input space to the output space. Before 1950's, control problems were largely treated as filtering problems and in the frequency domain, which is particularly suitable for single-input and single-output systems in classical control theory.

During 1950's Rodolf Emil Kalman puts forward a state space description for control systems. A set of state variables were introduced to describe the control system. Intuitively, the control system is divided into two parts: the first part is a set of differential or difference equations (called state equation) which are employed to describe the dynamics from input variable u to state variables x, and then a algebraic equation(called output equation) is used to describe the mapping from system state variable vector x to output vector y.

For specific control systems, there are different types of state space descriptions that could realize the same function description of control system. These state space models are called state space realization of input-output mapping. If there is no other realization that has less dimension of the state space model, such realization is called minimum realization of control system.

Feedback is possibly the most fundamental concept in control system and has a very

long history. Feedback means that the control strategy depends on the output vector and system state vector of system. Hence in modern control system, the feedback can be classified into state feedback control and output feedback control which have been in previous chapters of this book.

A nonlinear control system is always described by the following differential equation and output equation in time domain.

$$\dot{x} = f(x,\ u,\ t),\quad x \in R_x,\quad u \in U \tag{12.1a}$$

$$y = h(x,\ u,\ t),\quad y \in N \tag{12.1b}$$

Here, R_x, U, N are all kinds of dimensions n, m, l respectively. $f(x,\ u)$ and $h(x)$ are smooth mapping relation. In many applications we have $M = R^n,\ U = R^m,\ N = R^l$. When the expressions of $f(x,\ u,\ t)$ and $h(x)$ are both linear equation, equation set(12.1) represents a linear system. If they are nonlinear, equation(12.1) does a nonlinear system.

In the case of discrete systems, the following difference equations are used:

$$x(k+1) = f[k,\ x(k),\ u(k)] \tag{12.2a}$$

$$y(k) = h[k,\ x(k),\ u(k)] \tag{12.2b}$$

Here, x is $n\times 1$ state vector, $x \in R^n$; u is $m\times 1$ input vector, $u \in R^m$, and y is $l\times 1$ output vector, $y \in R^l$.

Reviewing the history of modern control theory, people know that the linear theory was firstly put forward and developed for linear systems. And then as the control theory for linear control systems had become very mature, control theory for nonlinear control systems started to be considered and studied. So in nonlinear control theory, a lot of concepts and results from linear systems are originated, inherited, and developed, or extended to nonlinear systems, however many results in linear system can not be applied to nonlinear systems, for example homogeneity does not apply. The response to an input $\alpha u(t)$ is not just α times the response to $u(t)$. The response to $\beta x(t_0)$ is not just β times the response to $x(t_0)$. The whole concept of designing control systems based on typical test inputs(unit steps, sinusoid, cosine, and so on) and then predicting behavior to an actual input by scaling and superposition is generally invalid.

Though the linear control theory is one of the major foundation of the nonlinear control theory, nonlinear control systems are more complicated than linear control systems, and furthermore has itself unique properties. In fact, linear systems are a particular and simplest class of nonlinear systems. *Dynamic performance of a nonlinear system depends on the system's parameters and initial conditions, as well as on the form and the magnitude of external actions.* The basic solution of the differential equation of nonlinear are generally very complex and various.

In this chapter, we will only coarsely narrate the main theories or opinions on

nonlinear control systems.

12.1.1 Basic problem of nonlinear control system

Control systems with at least one nonlinear element are called nonlinear systems. Their dynamics are described by nonlinear mathematical model. Contrary to linear system theory, *there does not exist a general theory applicable to all nonlinear systems*. Instead, a very complicated mathematical apparatus is employed, such as functional analysis, differential geometry or the theory of nonlinear differential equations. Except some special cases such as Riccati equations and equations with elliptic integral, a general solution can be difficultly found. Instead, individual procedures are applied, which are often improper and too complex for engineering practice. This is the reason why a series of approximate procedure are in use, in order to get some necessary knowledge about the system's dynamic properties. With such approximate procedures, nonlinear characteristics of real elements are substituted by idealized ones, which can be described mathematically.

Current procedures for the analysis of nonlinear control systems are classified into two categories: exact and approximate. Exact procedures are applied if approximate procedures do not yield satisfying results or when a theoretical basis for various synthesis approaches is needed.

The nonlinear control systems comprises two basic problems:

(1) *Analysis problem* consists of theoretical and experimental research in order to find either the properties of the system or an appropriate mathematical model of the system.

(2) *Synthesis problem* consists of determining the structure, parameters, and control system elements in order to obtain desired performance of a nonlinear control system. Further, a mathematical model must be set as well as the technical realization of the model. Since the controlled object is usually known, the synthesis consists of defining a controller in a broad sense.

12.1.2 Basic specific properties of nonlinear control system

Some common properties of nonlinear control system are given out in the following:

1.Unbounded reaction in finite time interval

The output signal of an unstable linear system increases beyond boundaries, when $t \to \infty$. With a nonlinear system, the output signal can increase indefinitely in finite time. For example, the output signal of the nonlinear system described by the differential equation of the first order $\dot{x} = x^2$ with initial condition $x(0) = x_0$ tends to infinity for $t = 1/x_0$.

2.Equilibrium state of nonlinear system

Linear stable systems have one equilibrium state. For example, the response of a linear

stable system to unit pulse input is damped towards zero.

Nonlinear stable systems may possess several equilibrium states, i.e., a possible equilibrium state is determined by a system's parameters, initial conditions and the magnitudes and forms of external excitation. Hence, equilibrium state of nonlinear system is very complicated.

Example12.1 Dependency of equilibrium state on initial conditions

The first-order nonlinear system described by $\dot{x}(t) = -x(t) + x^2(t)$ with initial conditions $x(0) = x_0$, has the solution(trajectory) given by

$$x(t) = \frac{x_0 e^{-t}}{1 - x_0 + x_0 e^{-t}}$$

Depending on initial conditions, the trajectory can end in one of two possible equilibrium states $x_e = 0$ and $x_e = 1$, as shown in figure12.1.

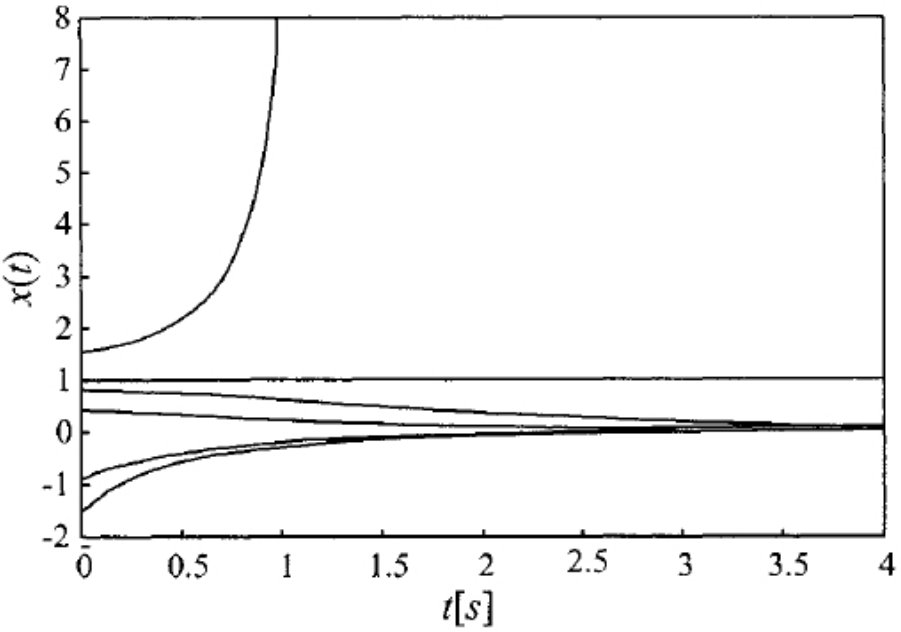


Figure12.1 State trajectories of a nonlinear system in example12.1

For all initial conditions $x(0) > 1$, the trajectories will diverge, while for $x(0) < 1$ they will approach the equilibrium state $x_e = 0$. For $x(0) = 1$ the trajectory will remain constant $x = 1$, $\forall t$, and the equilibrium state will be $x_e = 1$. Therefore, this nonlinear system has one stable equilibrium state $(x_e = 0)$ and one unstable equilibrium state, while the equilibrium state $x_e = 1$ can be declared as neutrally stable. The linearization o this nonlinear system for $|x| < 1$ (discarding nonlinear term) yields $\dot{x}(t) = -x(t)$ with the solution(state trajectory) $x(t) = x(0)e^{-t}$. This demonstrates that the unique equilibrium state $x_e = 0$ is stable for all initial conditions, since all the trajectories-notwithstanding initial conditions-will end at the origin.

3.Self-oscillations-Limit cycles

The possibility of oscillation without damp in a linear time-invariant system is linked with the existence of a pair of poles on the imaginary axis of the complex plane. The amplitude of oscillations is in this case given by initial conditions.

In nonlinear systems it is possible to have oscillation with amplitude and frequency which are not dependent upon the value of initial conditions, but their occurrence depends upon these initial conditions. Such oscillations are called *self-oscillations(limited cycles)* and they belongs to one of several stability concepts of the dynamic behavior of nonlinear systems.

4.Subharmonic harmonic or periodic and oscillatory process with harmonic inputs

In a stable linear system, a sinusoidal input causes a sinusoidal output with the same frequency. A nonlinear system under a sinusoidal input can produce an unexpected response. Depending on the type of nonlinearity, the output can be a signal with a frequency which is

- generally proportional to the input signal frequency,
- higher harmonic of the input signal frequency,
- -a periodic signal independent of the input signal frequency, or
- -a periodic signal with the same frequency as the input signal frequency.

With the input harmonic signal, the output signal can be a harmonic, subharmonic or periodic signal, depending upon the form, the amplitude and the frequency of the input signal.

5.Dynamic performance is not unique

In some cases, in a nonlinear system a pulse excitation provokes a response which tends to one or more stable equilibrium states under certain initial conditions.

6.Resonance jump

Resonance jump was investigated more in the theory of oscillation of mechanical system than in the theory of control systems. The term ”resonance jump” is used in case of a sudden jump of amplitude and/or phase and/or frequency of a periodic output signal of a nonlinear system. This happens due to nonunique relation which exists between periodic forcing input signal acting upon a nonlinear system and the output signal from that system. It is believed that resonance jump occurs in nonlinear control systems with small stability phase margin, i.e. With small damping factor of the linear part of the system and with amplitudes of excitation signal that force into the operating modes where nonlinear laws are valid, particularly saturation. Resonance jump can occur in nonlinear system operating in forced oscillation mode and is often not a desired state of the system. Resonance jump can not been seen from the transient response of the system and can not be defined by solving nonlinear differential equations. It is also not recommended to use experimental tests in plants during operation in order to resolve if system might have this phenomenon. To reduce or eliminate the resonance jump, higher stability phase margin is needed as well as the widening of the operating region of a nonlinear part of the system where the linear

laws are valid.

7.Synchronization

When the control signal is a sinusoidal signal of small amplitude, the output of a nonlinear system may contain subharmonic oscillations. With an increase of the control signal amplitude, it is possible that the output signal frequency"jumps" to the control signal frequency, i.e., synchronization of both input and output frequencies occurs.

8.Bifurcation

If there exist points in the space of system parameters where the system is not structurally stable, such points are called bifurcations points, since the system's performance"bifurcates". The values of parameters at which a qualitative change of the system's performance occurs are called critical or bifurcational values. The theory which encompasses the bifurcation problems is known as bifurcation theory(Gucken heimer and Holmes.1983).

If the bifurcation points exist in a nonlinear control system, it is very important to know the regions of structural stability in the parameter plane and in the phase plane and it is necessary to ensure that the parameters and the states of the system remain within these regions. If a bifurcation appears in nonlinear control system, the system can come to a chaotic state-this is of mostly theoretical interest since for safety reasons every control system has built-in activities which prevent any such situation. The consequence of bifurcation can be the transfer to a state with unbounded behavior. For the majority of technical systems this can lead to serious damages if the system has no built-in protection.

9.Chaos

In stable linear systems, small variations of initial conditions can result in small variations in response. Not so with nonlinear systems-small variations of initial conditions can cause large variations in response. This phenomenon is named *chaos*. It is a characteristic of chaos that the behavior of the system is unexpected, an entirely deterministic system which has no uncertainty in the modes of the system, excitation or initial conditions yields an unexpected response. Some mechanical systems as well as some electrical systems possess a chaotic behavior.

A chaotic system is one where trajectories present aperiodic behavior and are critically sensitive with respect to initial conditions. Here aperiodic behavior implies that the trajectories never settle down to fixed point or to periodic orbits. Sensitive dependence with respect to initial conditions means that very small differences in initial conditions can lead to trajectory that deviate exponentially rapidly from each other.

In nonlinear system, general functions have single-valued nonlinear functions $y = f(x)$, time-varying nonlinear functions $y = f(t, \ x)$ and multi-valued nonlinear characteristic functions which can be combined from elements with single-valued nonlinear characteristic and from the linear part of the system.

12.2 Linearization of Nonlinear Control System

12.2.1 Analysis on small deviation from nominal solution

Consider the general nonlinear state variable model in equation(12.1)

$$\dot{x} = f(x,\ \ u,\ \ t), \qquad y = h(x,\ \ u,\ \ t) \tag{12.3}$$

Suppose that a nominal solution $x_n(t)$, $u_n(t)$, and $y_n(t)$ on equation(12.3) is known. The difference between these nominal vector functions and some slightly perturbed functions $x(t)$, $u(t)$ and $y(t)$ can be defined by

$$\left.\begin{aligned}\delta x &= x(t) - x_n(t)\\ \delta u &= u(t) - u_n(t)\\ \delta y &= y(t) - y_n(t)\end{aligned}\right\} \tag{12.4}$$

Then equation(12.3) can be written as

$$\begin{aligned}&\dot{x}_n + \delta\dot{x} = f(x_n + \delta x,\ \ u_n + \delta u,\ \ t)\\ &= f(x_n,\ \ u_n,\ \ t) + \left[\frac{\partial f}{\partial x}\right]_n \delta x + \left[\frac{\partial f}{\partial u}\right]_n \delta u + higher-order\ \ terms\end{aligned} \tag{12.5a}$$

$$\begin{aligned}&y_n + \delta y = h(x_n + \delta x,\ \ u_n + \delta u,\ \ t)\\ &= h(x_n,\ \ u_n,\ \ t) + \left[\frac{\partial h}{\partial x}\right]_n \delta x + \left[\frac{\partial h}{\partial u}\right]_n \delta u + higher-order\ \ terms\end{aligned} \tag{12.5b}$$

Here $[\ ]_n$ means the derivatives are evaluated on the nominal solutions. Since the nominal solutions satisfy equation(12.3), the first terms in the preceding Taylor series exapansions cancel. For sufficiently small δx, δu and δy perturbations, the higher-order terms can be neglected, leaving the linear equations

$$\delta\dot{x} = \left[\frac{\partial f}{\partial x}\right]_n \delta x + \left[\frac{\partial f}{\partial u}\right]_n \delta u \tag{12.6a}$$

$$\delta y = \left[\frac{\partial h}{\partial x}\right]_n \delta x + \left[\frac{\partial h}{\partial u}\right]_n \delta u \tag{12.6b}$$

If $x_n(t) = x_e = const$ and if $u_n(t) = 0 = \delta u(t)$, then the stability of the equilibrium point x_e is governed by

$$\delta\dot{x} = \left[\frac{\partial f}{\partial x}\right]_n \delta x \tag{12.7}$$

For this case, the Jacobian matrix $[\partial f/\partial x]$ is constant matrix and its eigenvalues determine system stability in the neighborhood of x_e. If all eigenvalues λ_i have negative real parts, the equilibrium point is asymptotically stable for sufficiently small perturbations. If one or more eigenvalues have positive real parts, the equilibrium point is unstable. If one

or more of the eigenvalues are on the $j\omega$ axis and all others are in the left-half plane, no conclusion about stability can be drawn from this linear model. Whether the actual behavior of the system is divergent or convergent will depend upon the neglected higher-order terms in the Taylor series expansion. Thus, except for the borderline $j\omega$ axis case, stability of the nonlinear equation(12.3) is the same as the linearized model equation(12.6), at least in a small neighborhood of the equilibrium point.

Example12.2 Please find the equilibrium points for the system described by the following differential equation:

$$\ddot{y}+(1+y)\dot{y}-2y+0.5y^3=0$$

Then evaluate the linearized Jacobian matrix at each equilibrium point and determine the stability characteristics from the eigenvalues.

Solution

Letting $x_1=y$ and $x_2=\dot{y}$ gives the state variable equation

$$\dot{x}=\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}x_2\\ 2x_1-0.5x_1^3-(1+x_1)x_2\end{bmatrix}=f(x)$$

Equilibrium point are solutions of $f(x)=0$, hence have

$$x_2=0,\qquad 2x_1-0.5x_1^3=0$$

Three solutions(equilibrium points) of above equation set are

$$x_{e1}=\begin{bmatrix}0\\0\end{bmatrix},\quad x_{e2}=\begin{bmatrix}2\\0\end{bmatrix},\quad x_{e3}=\begin{bmatrix}-2\\0\end{bmatrix}$$

The Jacobian matrix is

$$\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2}\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2}\end{bmatrix}=\begin{bmatrix}0 & 1\\ 2-1.5x_1^2 & -1-x_1\end{bmatrix}$$

Therefore, for $x_{e1}=[0\quad 0]^T$, the Jacobian matrix is

$$\left.\frac{\partial f}{\partial x}\right|_{x=x_{e1}}=\begin{bmatrix}0 & 1\\ 2 & -1\end{bmatrix}$$

Its eigenvalues are $\lambda_1=1$ and $\lambda_2=-2$, so that this point is a saddle point.

For $x_{e2}=[2\quad 0]^T$, the Jacobian matrix is

$$\left.\frac{\partial f}{\partial x}\right|_{x=x_{e2}}=\begin{bmatrix}0 & 1\\ -4 & -3\end{bmatrix}$$

Its eigenvalues are $\lambda_{1,2}=\left(-3\pm j\sqrt{7}\right)/2$, so this point is a stable focus.

For $x_{e3}=[-2\quad 0]^T$, the Jacobian matrix is

$$\left.\frac{\partial f}{\partial x}\right|_{x=x_{e3}} = \begin{bmatrix} 0 & 1 \\ -4 & 1 \end{bmatrix}$$

Its eigenvalues are $\lambda_{1,2} = \frac{1 \pm j\sqrt{15}}{2}$, so this point is an unstable focus. If this system has initial conditions exactly at any one of the three equilibrium points, the state will remain there indefinitely in the absence of disturbances. For any other initial condition the state will eventually settle to $x = [2 \quad 0]^T$.

If a time-varying nominal solution $\{x_n(t), \quad u_n(t), \quad y_n(t)\}$ is used(perhaps as obtained from numerical solution of equation(12.3)), then the Jacobian matrices of equation(12.6) will be also time-varying in general. The stability of linear time-varying systems is not as straightforward as the linear, constant case.

With $\delta u(t)$ restricted to zero, equation(12.6) can be used to investigate the passive behavior of perturbed trajectories. It is of interest to know whether a trajectory $x(t)$ will passively return to $x_n(t)$(i.e., asymptotic stability) or will remain within some bounded neighborhood of it(i.e., stability) or will diverge from it(i.e.,unstable). These types of analysis must always be used with caution because of the assumptions made regarding $\delta x(t)$ remaining small.

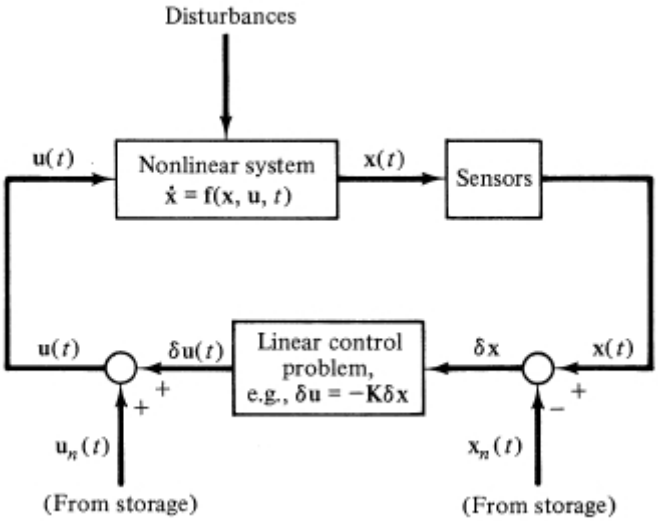


Figure12.2 A typical implementation of small deviation of system variable x

The input perturbation $\delta u(t)$ can be used to actively control the behavior of $\delta x(t)$, thus force it to return to and remain at or near zero. Then the linearizing assumption that $\delta x(t)$ is small can be made somewhat self-fulfilling. A linear feedback control law, such as $\delta u(t) = -K\delta x(t)$, can be used, and a typical implementation is shown in figure12.2. The overall goal is to maintain the trajectory near the known, pre-computed nominal in spite of initial condition perturbations or input disturbances. The gain matrix K in the control law could be computed using pole assignment techniques(see section6.4). If the closed-loop

poles are forced to be sufficiently stable, then $\delta x(t)$ will rapidly return to zero after any upset. Alternatively, the gain matrix K could be found as the result of an optimal regulator design problem.

12.2.2 Dynamic linearization using state feedback

In section12.2.1 we discuss the local linearization of a nonlinear system. In this section, we consider the problem of synthesizing a control input $u(t)$, which will cause the following system to have a response which matches some specified template system.

$$\dot{x} = f(x,\ \ u,\ \ t) \tag{12.8}$$

That's to say, let $y(t) = Cx(t)$. It is desired that system output response $y(t)$ matches as closely possible as the response of the specified template system

$$\dot{y}_d = g(y_d,\ \ y,\ \ t) \tag{12.9}$$

In a typical example, the g function can specify a linear system with $v(t)$ being perhaps a step function input and with the response possessing certain desirable transient characteristics.

$$\dot{y}_d = Fy_d + Gv \tag{12.10}$$

If a control input $u(t)$ can be found to achieve the goal $\dot{y} \equiv \dot{y}_d$, then the original system will behave as a linear system. This is what we term *dynamic linearization*.

Define the error $e(t) = Cx(t) - y_d(t)$, and then

$$\dot{e}(t) = C\dot{x}(t) - \dot{y}_d(t) = Cf(x,\ \ u,\ \ t) - g(y_d,\ \ v,\ \ t) \tag{12.11}$$

Assume that $C = E_{unit}$ (unit matrix). When $\dot{e}$ is set to zero, it may be possible to solve the resulting equation for the unknown input u in terms of known or measurable quantities x, y_d and v. If this is accomplished, the feedback-modified system will have the same derivative as the template system. If the template system is linear, then the original system will have been linearized. In nature, the nonlinearities of the original system are canceled and replaced by the desired linear terms. This form of dynamic linearization has been known for many years.

Example12.3 It is desired that the first-order nonlinear system

$$\dot{x} = x + u + xu$$

It behaves like the following linear system with initial condition $y_d(0) = 10$

$$\dot{y}_d = -\sigma y_d$$

Solution

For this scalar system let $y(t)=x(t)$ (namely error is zero). Setting $\dot{x}=\dot{y}_d$ leads to

$$u(t)=\frac{-x(t)-\sigma y_d(t)}{1+x(t)}$$

provided $x(t)\neq -1$.

At least two potential problems exist with this scheme:

(1) even if the derivatives can be made to match exactly, the initial condition $x(0)$ may not match $y_d(0)$ for a variety of reasons. The exact initial conditions may not be known due to measurement error, or the desire may be to have the system respond like the template system regardless of its initial value $x(0)$. Of course, matching derivatives does not mean matching response curves. This will be addressed in the sequel.

(2) The resulting control law for $u(t)$ has a singularity at $x(t)=-1$. An infinite amount of control would be required at this point. Forcing a nonlinear system to respond like a linear system generally means that components must be overdesigned to allow the avoidance of nonlinear behavior. The singularity of this example is an extreme case.

In the following we will discuss the problem of initial condition mismatch, either deliberate or unintentional which can be addressed by adding a convergence factor matrix S. Instead of setting $\dot{e}=0$, we require that

$$\dot{e}=Se \tag{12.12}$$

In a similar manner to the development for state observers in Chapter6, the matrix S is specified with asymptotically stable eigenvalues. Then $e(t)\to 0$, and thus $Cx(t)\to y_d(t)$ at a rate controlled by choice of convergence factor matrix S. The equation for finding the control $u(t)$ is thus

$$Cf(x,\ u,\ t)=g(y_d,\ v,\ t)+S[Cx(t)-y_d(t)] \tag{12.13}$$

The existence of a solution of equation(12.13) for $u(t)$ can be established in certain cases by using the implicit function theorem, which establishes sufficient conditions on the function $f(x,\ u,\ t)$. The solvability is also influenced by the number of independent equations that need to be satisfied relative to the number of unknown control components. This explains the existence of the matrix C. It will not generally be possible to match all n components of state variable x to an n-dimensional y_d vector when there are only $r<n$ components in the control vector u. For any specific problem, a direct attempt to solve for u will often be the most expedient method of determining whether or not such a solution can be found, and this is the approach presented in the example problems. Assuming the existence of a solution, the control law will be of the feedback form

$$u(t)=u(x(t),\ y_d(t),\ v(t),\ C,\ S) \tag{12.14}$$

And the procedure is obviously a model-matching or mode-tracking scheme, as shown in figure12.3.

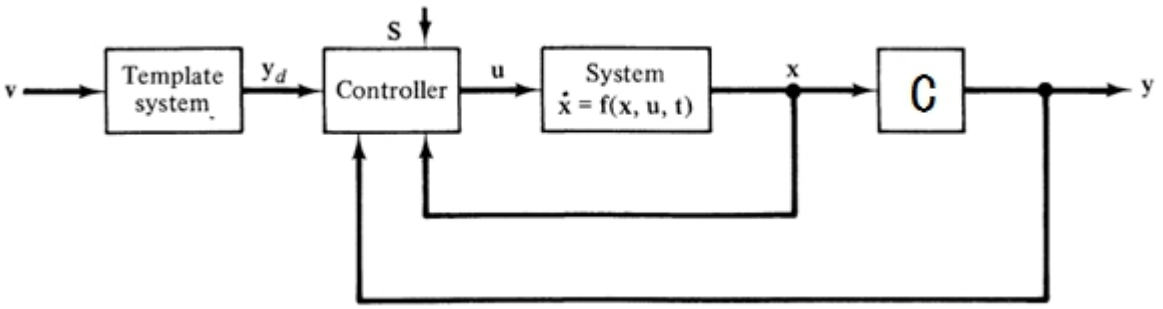


Figure12.3 Model-matching block diagram

Note that when a linear system is used as the template, then equation(23.14) becomes

$$Cf(x,\ \ u,\ \ t) = Fy_d + Gv + S(Cx - y_d) = (F - S)y_d + SCx + Gv \quad (12.15)$$

If the convergence matrix S is selected equal to F, then y_d is not directly required, and the need to synthesize the template system is removed. If matrix $S = 0$ is selected, a major feedback path is eliminated. Then equation(12.15) becomes

$$Cf(x,\ \ u,\ \ t) = Fy_d + Gv \quad (12.16a)$$

When $S = F$, equation(12.15) becomes

$$Cf(x,\ \ u,\ \ t) = SCx + Gv = FCx + Gv \quad (12.16b)$$

Example12.4 The scalar system of example12.3 is reconsidered, but now the convergence factor matrix S is included so that for all initial conditions, $x(0)$, $x(t)$ will ultimately approach the desired response $\dot{e} = \dot{x} - \dot{y}_d = Se$, where S is a negative real number. Then

$$x + u + xu + \sigma y_d = S(x - y_d)$$

From which, if $x(t) \neq -1$,

$$u(t) = \frac{S[x(t) - y_d(t)] - x(t) - \sigma y_d}{1 + x(t)}$$

Substituting this back into the system equation gives the coupled pair

$$\begin{bmatrix} \dot{x} \\ \dot{y}_d \end{bmatrix} = \begin{bmatrix} S & -(\sigma + S) \\ 0 & -\sigma \end{bmatrix} \begin{bmatrix} x \\ y_d \end{bmatrix}$$

The above 2×2 transition matrix is easily found:

$$\phi(t,\ \ 0) = \begin{bmatrix} e^{St} & e^{-\sigma t} - e^{St} \\ 0 & e^{-\sigma t} \end{bmatrix}$$

Then the system output response is

$$x(t) = e^{-\sigma t} y_d(0) + [x(0) - y_d(0)]e^{St}$$

This is the desired template response, $e^{-\sigma t} y_d(0)$, plus an initial condition mismatch term, which dies out at a rate determined by the convergence factor matrix S. If $S \leq -\sigma$, this term will quickly die out, leaving the desired response.

Besides small deviation and state feedback methods, we can also utilize the describing function method to describe the nonlinear system. Other linearization methods in nonlinear system have graphical linearization, harmonic linearization, statistical linearization, least square, and so on. If readers are interested in these given methods, readers may consult related control books.

12.3 Stability, Controllability and Observability

Stability, controllability and observability have been narrated to analyze the state of an linear system in previous chapters. In Chapter4, many contents on stability can be applied into nonlinear system such as Lyapunov stability criterion and so on. Because present nonlinear theory is not mature at present, we only coarsely demonstrate related topics th

12.3.1 Nonlinear system stability

1. Stability of a nonlinear system based on stability of the linearized system

Stability of a nonlinear system can be analyzed through the stability of a linearized one, but caution must be taken. The linking of the equilibrium states of a linearized system and the original nonlinear system can be seen in an example of an unforced system of second order with equations.

$$\left.\begin{aligned}\dot{x}_1 &= f_1(x_1,\ \ x_2)\\ \dot{x}_2 &= f_2(x_1,\ \ x_2)\end{aligned}\right\} \tag{12.17}$$

Linearizing the nonlinear system in the vicinity of the equilibrium state, and considering the behavior of the obtained linear system, it is possible—except in specific situations — to analyze the behavior of the nonlinear system in the vicinity of the equilibrium state. Supposing that the equilibrium state of the nonlinear system by equation(12.17) is at the origin , and that f_1 and f_2 are continuously differential near the o $[x_{e1}\quad x_{e2}]^T=[0\quad 0]^T$ origin (equilibrium point), then the Taylor expansion near the origin gives

$$\begin{aligned}\dot{x}_1(t) &= f_1(x_1,\ \ x_2)=f_1(0,\ \ 0)+a_{11}x_1+a_{12}x_2+r_1(x_1,\ \ x_2)\\ &= a_{11}x_1+a_{12}x_2+r_1(x_1,\ \ x_2)\end{aligned} \tag{12.18a}$$

$$\begin{aligned}\dot{x}_2(t) &= f_2(x_1,\ \ x_2)=f_2(0,\ \ 0)+a_{21}x_1+a_{22}x_2+r_2(x_1,\ \ x_2)\\ &= a_{21}x_1+a_{22}x_2+r_2(x_1,\ \ x_2)\end{aligned} \tag{12.18b}$$

Here, $r_1(x_1,\ \ x_2)$ and $r_2(x_1,\ \ x_2)$ are higher-order terms of the Taylor series or remainders. As the equilibrium point at the origin is $f_1(0,\ \ 0)=0$ and $f_2(0,\ \ 0)=0$, the linearized model follows:

$$\dot{z}_1 = a_{11}z_1 + a_{12}z_2 \tag{12.19a}$$

$$\dot{z}_2 = a_{21}z_1 + a_{22}z_2 \tag{12.19b}$$

Or

$$\dot{z}(t) = Az(t) \tag{12.20}$$

Here:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad a_{ij} = \left[\frac{\partial f_i}{\partial x_j}\right]_{x=0}, \quad i,j = 1,2; \quad z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^T$$

The analytical procedure of linearization is based on the fact that if the matrix A has no eigenvalue with $R_e(\lambda_i) = 0$, the trajectories of the nonlinear system(12.17) in the vicinity of the equilibrium state $\begin{bmatrix} x_{e1} & x_{e2} \end{bmatrix}^T = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ have the same form as the trajectories of the linear system(12.19) in vicinity of the equilibrium state $\begin{bmatrix} z_{e1} & z_{e2} \end{bmatrix}^T = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$. Table2.1 shows the types of equilibrium states(points) that are determined from the singular points.

Table12.1: Types of singular points of a nonlinear and a linearized system

Eigenvalues(λ_i) of linear system(2.19)	Equilibrium state $\begin{bmatrix} z_{e1} & z_{e2} \end{bmatrix}^T$ of linear system(2.19)	Equilibrium state $\begin{bmatrix} z_{e1} & z_{e2} \end{bmatrix}^T$ of nonlinear system(2.17)
Real and negative	Stable node	Stable node
Real and positive	Unstable node	Unstable node
Real of opposite sign	Saddle	Saddle
Conjugate complex with negative real part	Stable focus	Stable focus
Conjugate complex with positive real part	Unstable focus	Unstable focus
Imaginary(single)	Center	Undefined

If the equilibrium state of the linearized mathematical model(12.19) is of the type center, then the linearized system oscillates with constant amplitude. The behavior of the trajectory of the original nonlinear system(2.17) is determined by the remainder of Taylor series $r_1(x_1, x_2)$ and $r_2(x_1, x_2)$ that were neglected during the linearization process. Analysis of the linearized system alone gives in this case no final answer about the behavior of the nonlinear system. In order to clarify this situation, the following example is given here.

Example12.5 A oscillator is described by the following nonlinear differential equation:

$$\ddot{y} - \mu(1 - y^2)\dot{y} + y(t) = 0; \quad \mu = const > 0$$

After choosing the state variables $x_1 = y$, $x_2 = \dot{y}$ the state-space equation is

$$\left.\begin{aligned}\dot{x}_1(t) &= x_2 \\ \dot{x}_2(t) &= -x_1 + \mu(1-x_1^2)x_2\end{aligned}\right\}$$

Linearization at the equilibrium point $[x_{e1} \quad x_{e2}]^T = [0 \quad 0]^T$ gives

$$\left.\begin{aligned}\dot{z}_1(t) &= z_2 \\ \dot{z}_2(t) &= -z_1 + \mu z_2\end{aligned}\right\}$$

And linearized matrix A :

$$A = \begin{bmatrix} 0 & 1 \\ -1 & \mu \end{bmatrix}$$

Eigenvalues of the matrix can be gained from the characteristic equation:

$$\lambda^2 - \lambda\mu + 1 = 0, \quad \lambda_{1,2} = \frac{\mu \pm \sqrt{\mu^2 - 4}}{2}$$

If $\mu > 0$, the eigenvalues have a positive real part, and the equilibrium point of the linearized mathematical model $[z_{e1} \quad z_{e2}]^T = [0 \quad 0]^T$ is of the unstable focus. The original nonlinear system will have an equilibrium state ate the origin $[z_{e1} \quad z_{e2}]^T = [0 \quad 0]^T$ of the unstable focus.

2. Absolute stability of equilibrium state of unforced system

Absolute stability of nonlinear control systems was firstly published in 1944. Lur'e and Postinkov had researched nonlinear systems with a continuous single-valued nonlinear characteristic which passes through the first and third quadrants in 1944. M.A. Aizerman formulated the problem of absolute stability when the nonlinear characteristic is within a sector-in 1947 the established a hypothesis whereby the stability of nonlinear system may be analyzed with linear procedures. The Romanian mathematician V.M. Popov in 1959 proposed a fundamentally new approach to the problems of absolute stability — he established necessary conditions which the amplitude-frequency characteristic of the linear part of the system must fulfill, so that the nonlinear system will be absolutely stable. This frequency immediately found the approval of engineers to whom the frequency approach was familiar. In fact, the concept of absolute stability means a global asymptotic stability of the equilibrium states in the Lyapunov sense.

The structure of the nonlinear system in problems of absolute stability is given in figure12.4. In the direct branch is the linear time-invariant system, while in the feedback branch is a single-valued nonlinearity (nonlinear element without memory), which means that the feedback performs nonlinear static mapping of the signal e_2 to the signal y_N. The signal r_1 may represent a reference signal, while the signal r_2 may represent an error signal, for example measurement noise. In the case that the transfer function of the linear part is strictly proper $(D = 0)$ and $r_1 = r_2 = 0$, the system in figure12.3 can be

mathematically described by

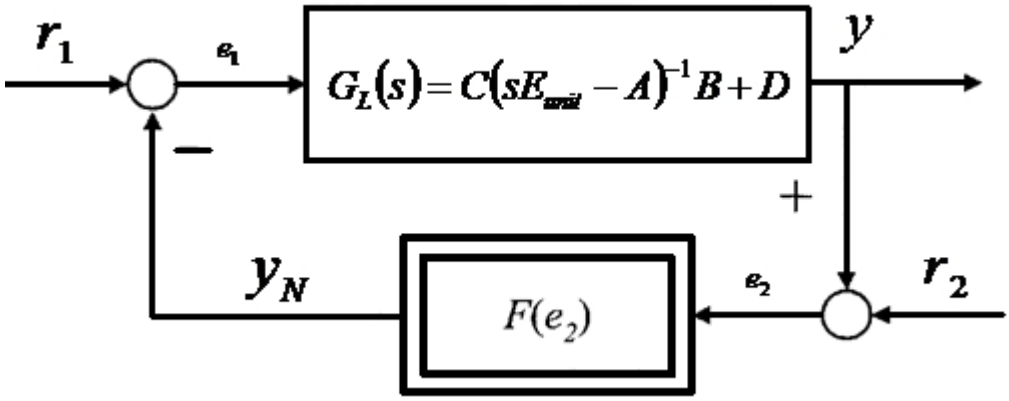


Figure12.4 Structure diagram of nonlinear system in the analysis of absolute stability

$$\dot{x} = Ax - BF(y) \tag{12.21a}$$

$$y = Cx \tag{12.21b}$$

In equation(12.21), $y = e_2$ and $e_1 = -y_N = -F(y)$. Many automatic control systems can be represented with this structure. If the system acts in a stabilization mode(which is in nature a regulator problem), the structure from figure12.4 can be simplified to the structure in figure12.5. If the nonlinear function $F(y)$ belongs to the sector $[k_1, \ k_2]$ and if the linear part of the system is stable, people possibly ask what additional condition is necessary for nonlinear system?

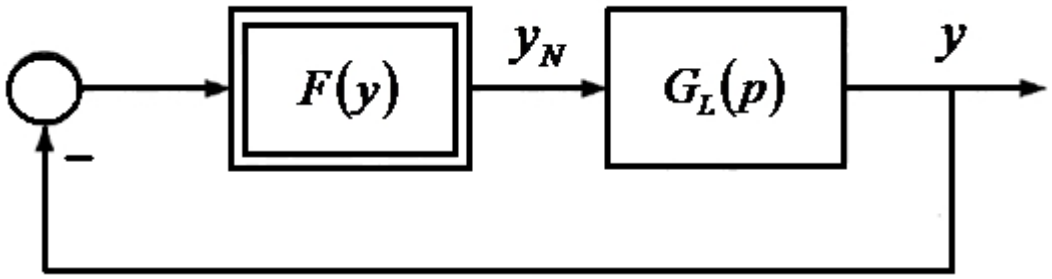


Figure12.5 Structure diagram of unforced nonlinear system

As the nonlinear characteristic is in the sector $[k_1, \ k_2]$, that means it is bounded by two straight line which intersect at the origin—this corresponds to feedback with constant gain—it is normal to suppose that the stability of the nonlinear system will have certain similarities with the stability of the system which is stabilized by the gain in the feedback loop. Contrary to the stability analysis at present, people's interest is not in a specific system, but in the whole family of system, as $F(y)$ can be any nonlinear function inside the sector $[k_1, \ k_2]$. This is the reason why this is called the problem of absolute stability in the sense that if the system is absolutely stable, it is stable for the whole family of nonlinearities in the sector $[k_1, \ k_2]$.

In 1949, M.A.Aizerman considered this problem and established the following hypothesis. If the system described by

$$G_L(s)=\frac{B(s)}{A(s)}=\frac{K_L\left(b_m s^m+b_{m-1}s^{m-1}+\cdots+1\right)}{a_n s^n+a_{n-1}s^{n-1}+\cdots+1};\quad m\le n-1 \tag{12.22}$$

then the system in equation(12.22) is globally asymptotically stable for all linear mappings given by

$$F(t,\quad y)=ky,\quad \forall t,y\quad k\in[k_1,\quad k_2] \tag{12.23}$$

Here, the gain k of the linear feedback is inside the interval $[k_1,\quad k_2]$, then the same is true for all time-invariant nonlinear system with a single-valued static characteristic $F(y)$ inside the sector $[k_1,\quad k_2]$. In other words, if the nonlinear feedback is replaced by a linear proportional feedback and if we obtain a closed-loop system globally asymptotically stable for all values of the linear gain inside $[k_1,\quad k_2]$, then the nonlinear system which possesses a nonlinear feedback is globally asymptotically stable if the static characteristic of the nonlinear element is inside the sector $[k_1,\quad k_2]$. This attractive hypothesis is not valid in the general case. R.E.Kalman(1957) has proposed a similar hypothesis which assumes that $F(y)$ belongs to the incremental sector $[k_1,\quad k_2]$. The set of nonlinearities in the incremental sector is smaller than the set of nonlinearities treated by Aizerman so the Kalman hypothesis has greater probability to be correct.

Nonlinear systems with the structure shown in figure12.6 will be considered further. The linear part of the system can be stable, unstable or neutrally(critically) stable, with the transfer function expressed in equation(12.22).

Nonlinear part may be the following cases

(1) Single-valued time-invariant nonlinear element $y_N(t)=F[x(t)]$,

(2) Single-valued time-varying nonlinear element $y_N(t)=F[t,\quad x(t)]$,

(3) Double valued time-invariant nonlinear element $y_N(t)=F\left[x(t),\quad \dot{x}(t)\right]$

Besides the above-mentioned nonlinear elements, there appear:

(4) Linear time-varying element $y_N=k(t)x(t)$,

(5) Linear time-invariant element $y_N=kx(t)$.

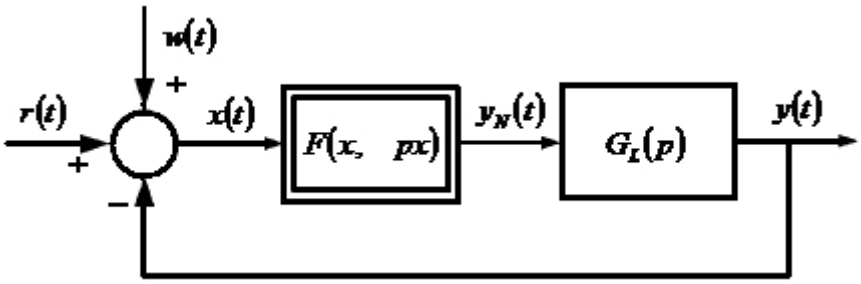


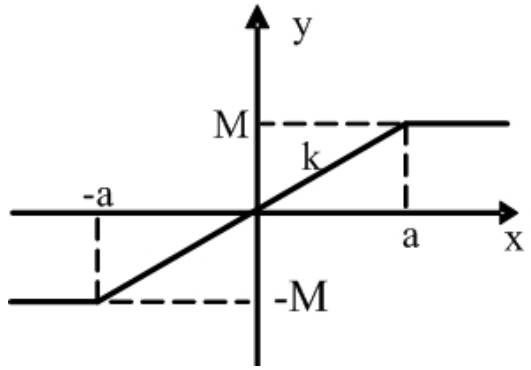
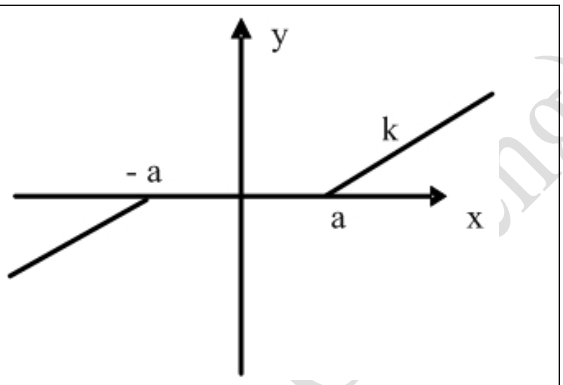
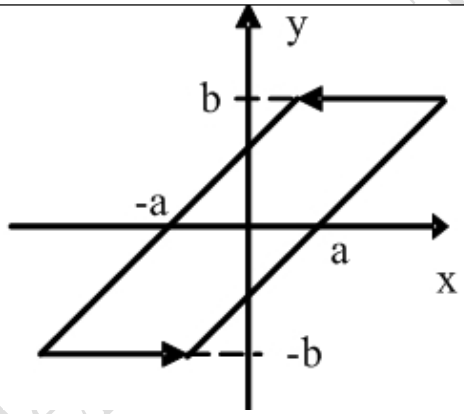
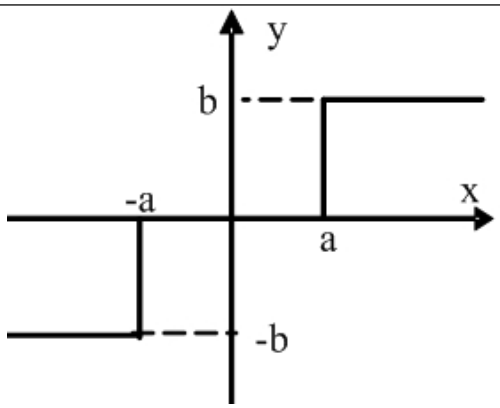
Figure12.6 Basic structure of a nonlinear system for analysis of absolute stability with external signals $r(t)$ and $w(t)$

All these elements must have static characteristic inside the sector $[k_1,\quad k_2]$ for all

$t > 0$, where $0 < k_1 < k_2 < \infty$.

In closed-loop control system, the classical nonlinear characteristics are listed in table12.2 to aid readers to understand common nonlinear element of nonlinear system.

Table12.2 Classical nonlinear characteristics in closed-loop control system

<table>
<tr>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td>1. Saturation characteristic
Math description:
$$y=\begin{cases} M & x > a \\ kx & |x| < a \\ -M & x < -a \end{cases}$$
Characteristic: when input signal is very small, element work in a linear area; when input signal exceeds the linear area, system output takes on saturation state.</td>
<td>2. Dead zone characteristic
Math description
$$y=\begin{cases} 0 & |x| \le a \\ k(x-a) & x > a \\ k(x+a) & x < -a \end{cases}$$
Characteristic: when input signal is very small, there is no output signal; when input signal is over the range of dead zone, system output changes as the change of input signal.</td>
</tr>
<tr>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td>3. Hysteresis(backlash) characteristic
Math description
$$y=\begin{cases} k(x-a\,\mathrm{sign}\dot{x}) & \dot{y} \neq 0 \\ b \cdot \mathrm{sign}x & \dot{y} = 0 \end{cases}$$
Characteristic: when input signal is very small, output is zero; when input is greater than a, output is proportional to input. When input is inverse, output keeps constant; output is not linear change until $r(t) < -a$.</td>
<td>4. Relay characteristic
Math description
$$y=\begin{cases} b & x > a \\ -b & x < -a \end{cases}$$
Characteristic: double-position, dead-zone, backlash characteristic. Phase-plane and describing function methods can be utilized to analyze the affect of relay nonlinear characteristic to system performance.</td>
</tr>
</table>

The reference signal $r(t)$ and disturbance signal $w(t)$ act on the closed-loop nonlinear system. The total input signal to the system is therefore $f(t)=r(t)+w(t)$. The dynamics of the closed-loop nonlinear system in figure12.6 can be described by the differential equation:

$$x(t)=f(t)-G_L(p)F(x) \tag{12.24}$$

Or by the following integral equation

$$x(t)=f(t)-\int_0^t g(t-\tau)F(x)d\tau \tag{12.25}$$

Here $g(t)$ is the weighting function of the linear part of the system, $f(t)$ is the total external signal acting on the control system, $F(x)$ is a mathematical description of nonlinear element of the system and $p=d/dt$ is a derivative operator.

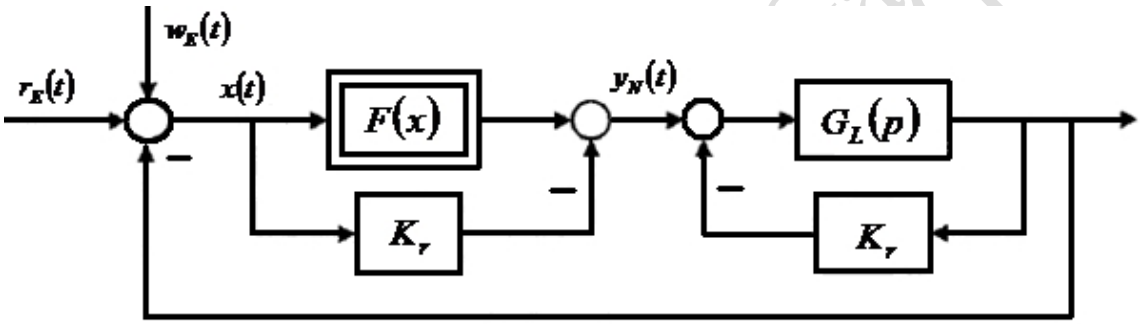


Figure12.7 Equivalent structure of nonlinear system in the case of unstable linear part

Absolute stability of the equilibrium of the system in figure12.6 is best treated by the frequency criterion of V.M. Popov who gave out the judgment criterion. If you want to apply Popov's criterion to nonlinear control system, this system must be time-invariant and, moreover, the linear part of the system $G_L(s)$ is stable. In the case when the linear part is unstable, the system's structure is replaced by equivalent structure shown in figure12.7. As is evident from the block diagram, the unstable linear part of the system is stabilized with the linear operator K_r in the negative feedback loop. The same linear operator is placed in parallel with the nonlinear element, so that the influence of the stabilized feedback on the dynamics of the closed-loop system is eliminated. Equivalent external actions(reference input and disturbance), equivalent nonlinear element and equivalent linear part of the system are given by the following expression:

$$f_E(t)=r_E(t)+w_E(t)=\frac{R(p)}{1+K_rG_L(p)}+\frac{W(p)}{1+K_rG_L(p)} \tag{12.26a}$$

$$G_E(p)=\frac{G_L(p)}{1+K_rG_L(p)} \tag{12.26b}$$

$$F_E(x)=F(x)-K_rx \tag{12.26c}$$

Here, K_r is a stable linear operator which stabilizes the otherwise unstable linear part of the system $G_L(s)$.

In the following, systems with a stable linear part in figure12.6 and systems with equivalent structure in figure12.7 will be discussed. The external actions on the system $f(t)=r(t)+w(t)$ can be divided into two groups:

(1) Bounded external actions $f_1(t)$ described by the following equations:

$$|f_1(t)|<M_1;\quad t\geq 0 \tag{12.27}$$

This function can represent reference and other disturbance inputs.

(2) Vanishing(time-decreasing) external actions $f_2(t)$ described by the following expressions:

$$\int_0^\infty |f_2(t)|dt<M_2;\quad t\geq 0 \tag{12.28a}$$

$$\lim_{t\to\infty} f_2(t)=0 \tag{12.28b}$$

which represent initial conditions different from zero.

Dynamics of the system in figure12.6 which was at rest up to the moment $t=0$, when the external signal $f(t)=f_1(t)+f_2(t)$ was applied, are described by the following integral expression

$$x(t)=f_1(t)+f_2(t)-\int_0^t g(t-\tau)F(x)d\tau \tag{12.29}$$

In order to find the absolute stability of the solution(12.29), it is appropriate to look separately at the equilibrium states of the forced system($f_1(t)\neq 0$) and those of the unforced one($f_1(t)\neq 0$ and $f_2(t)\neq 0$).

In the case when only a vanishing external action $f_2(t)$ acts on the nonlinear system, the absolute stability(global asymptotic stability) of the equilibrium state x_e of the unforced system is considered. The solution(a zero-input response) $x_{zi}(t)$ will be asymptotically stable if

$$x_e=\lim_{t\to\infty} x_{zi}(t)=M_x=const \tag{12.30a}$$

Or

$$x_e=\lim_{t\to\infty} x_{zi}(t)=0 \tag{12.30b}$$

It must be stressed here that in the case when a stable nonlinear system has the nonlinearity of the type dead zone in table12.2, its equilibrium state $x_e=x_{zi}(\infty)=M_x\leq|x_a|$ may belong to any part of the dead zone(part of the stability), i.e., the nonlinear system can possess an infinite number of equilibrium states, so the condition of asymptotic stability(12.30) can not be applied. Therefore it is more appropriate to consider the equilibrium state as stable if the following condition is satisfied:

$$\lim_{t\to\infty}|x_{zi}(t)-x_e|=0 \tag{12.31}$$

Here, $x_e=x_{zi}(\infty)=M_x$ is any value inside the dead zone $-x_a<M_x<x_a$. In accordance with the definition of asymptotic stability(12.31), we distinguish the local

asymptotic stability—when condition(12.31) is satisfied for small deviations $f_2(t)$ from the equilibrium state, and global asymptotic stability — when the condition(12.31) is satisfied for large deviations $f_2(t)$ from the equilibrium state.

Contrary to linear systems where local asymptotic stability assures global asymptotic stability, in nonlinear systems local asymptotic stability may exist, but not a global one.

Generally, two approaches to the problem of stability are possible. The first method is to find *the solution of the differential(12.24) or integral(12.25) equation*, which is in practice not applicable because of well-known difficulties. The second one is to *determine the stability conditions without the inevitable quest for the solution of the dynamic equations of the system*. This approach is necessary because of the fact that quite often the nonlinear characteristic $y_N = F(x)$ can not be determined. Namely, the dynamics of the nonlinear system are changing with the change of operating conditions. For example, change of the load or of the supply energy of the control mechanism results in deformity of the static characteristic, which greatly complicates the exact determination of values of the parameters of a differential or integral equation of the nonlinear system.

The static characteristic of many actuators of modern control systems can be regarded as nonlinear functions with the following properties:

$$xF(x) > 0, \quad x \neq 0 \tag{12.32a}$$

$$F(0) = 0 \tag{12.32b}$$

Here, $F(x)$ is a continuous function:

$$\int_0^{\pm\infty} F(x)dx = \pm\infty \tag{12.33}$$

Nonlinear functions with the properties of equations(12.32) and (12.33) can have very different graphical presentations. For nonlinear system given in figure12.6,the characteristic of the nonlinear element is situated within the sector bounded by $k_1 x$ and $k_2 x$:

$$k_1 < \frac{F(x)}{x} < k_2; \quad x \neq 0 \tag{12.34}$$

If the nonlinear function $F(x)$ is located in the sector $[k_1, \ k_2]$ and fulfills the conditions(12.32) and (12.33), the global asymptotic stability of the system with the function $y_N = F(x)$ is called absolute stability. Very often the nonlinear functions can be of the class $[0, \ k_2]$ and $[0 \ \infty]$ which result from(12.32) for $k_1 = 0$ and for $k_1 = 0$, $k_2 = \infty$.

3. Absolute stability of equilibrium state of unforced nonlinear system

The Romanian mathematician V.M.Popov formulated in 1959 frequency criterion of the absolute stability of a time-invariant unforced nonlinear system which has the structure in figure12.6. With $f(t) = w(t), \ r(t) = 0$ a system is described, and a vanishing external quantity $w(t) = f_2(t)$ (initial condition) which satisfies conditions (12.28a) and (12.28b) is

applied. The time-invariant linear part of the system has a stable equilibrium state, while the nonlinear functions of the class $[0,\ \ k_2]$ that satisfy conditions (12.32) and (12.33). It is:

$$F(0)=0 \tag{12.35a}$$

$$xF(x)>0;\quad x\neq 0 \tag{12.35b}$$

$$\int_0^{\pm\infty} F(x)dx=\pm\infty \tag{12.35c}$$

$$k_1<\frac{F(x)}{x}<k_2;\quad x\neq 0 \tag{12.35d}$$

The Popov criterion of absolute stability for an unforced nonlinear system which has only the vanishing external quantity $f_2(t)$, or the initial condition which differs from zero, and shown in the block diagram in figure12.6, is formulated as follows:

Theorem12.3.1: Popov criterion of absolute stability— G_L stable

The equilibrium state of an unforced nonlinear control system of the structure as in figure12.6 will be globally asymptotically stable—absolutely stable if the following is true:

(1) Linear part of the system is time-invariant, stable and completely controllable.

(2) Nonlinear function $F(x)$ is of class $[0,\ \ k_2]$ with $0<k_2<+\infty$ and satisfies condition(12.35a)~(12.35d).

(3) There exist two strictly positive real numbers $q>0$ and an arbitrarily small number $\delta>0$, such that for all $\omega\geq 0$ the following inequality is true(Popov, 1973):

$$\mathrm{Re}\{(1+jq\omega)G_L(j\omega)\}+\frac{1}{k_2}\geq\delta>0 \tag{12.36}$$

Or,

$$\mathrm{Re}\{(1+jq\omega)G_L(j\omega)\}+\frac{1}{k_2}>0 \tag{12.37}$$

Here,

$$k_2<\infty;\quad \lim_{\omega\to\infty}G_L(j\omega)=0 \tag{12.38}$$

The Popov criterion enables the relatively simple determination of the stability of the nonlinear system, based on knowledge of the sector where eventually the nonlinear static characteristic lies and based on knowledge of the frequency characteristic of the linear part of the system. There are special cases which allow the linear part to have one or two poles at the origin. In such cases—besides inequality(12.36) and (12.37)—the following must be used:

(1) When $G_L(s)$ has one pole at the origin

$$\lim_{\omega\to+0}\{\mathrm{Im}|G_L(j\omega)|\}\to-\infty \tag{12.39}$$

(2) When $G_L(s)$ has two poles at the origin:

$$\lim_{\omega\to+0}\{\mathrm{Re}|G_L(j\omega)|\}\to-\infty \tag{12.40}$$

For small ω, $\mathrm{Im}\{G_L(j\omega)\}<0$.

The inequality(12.36) or (12.37) is called the Popov inequality. Here it must be emphasized that the Popov criterion gives only the sufficient condition for stability. Its importance lies in the fact that the stability of the nonlinear system can be evaluated on the basis of the frequency characteristic of the linear part of the system, without the need for seeking a Lyapunov function. The criterion is constrained by the requirement that the nonlinear static characteristic must be single-valued and that it lies in the first and third quadrants $(k_1>0)$—it must pass through the origin.

When the nonlinear element $y_N(t)=F\left[x(t),\ \dot{x}(t)\right]$ is single-valued and time-varying, it is necessary to put $q=0$ into equation(12.37). For double-valued time-invariant nonlinear elements $y_N(t)=F\left[x(t),\ \dot{x}(t)\right]$, the value $q=0$ in equation(12.37) is also used.

In the analysis and the synthesis of nonlinear control systems of the proposed structure in figure12.6, the most appropriate procedure is the geometric interpretation of the criterion of absolute stability as it enables the treatment of nonlinear systems by applying frequency methods which were developed in the theory of linear control systems.

In order to determine q which satisfies criterion(12.37), V.M.Popov has proposed a geometrical interpretation of the analytic condition, so that instead of the frequency characteristic of the linear part of the system $G_L(j\omega)$, a modified frequency characteristic $G_p(j\omega)$—the Popov characteristic or Popov polt—is used.

$$G_p(j\omega)=\mathrm{Re}\{G_L(j\omega)\}+j\omega\,\mathrm{Im}\{G_L(j\omega)\}=U(\omega)+jV_p(\omega) \tag{12.41}$$

Here, $V_p(j\omega)=\omega V(\omega)$ is the imaginary part of the Popov characteristic.

Replacing equation(12.41) into equation(12.37) gives the criterion of absolute stability which include the Popov plot $G_p(j\omega)$:

$$\mathrm{Re}\{G_p(j\omega)\}-q\,\mathrm{Im}\{G_p(j\omega)\}+\frac{1}{k_2}>0 \tag{12.42}$$

Or

$$U(\omega)-q\omega V(\omega)+\frac{1}{k_2}>0 \tag{12.43}$$

The boundary value(12.43) is the equation of the straight line-Popov line:

$$U(\omega)=q\omega V(\omega)-\frac{1}{k_2} \tag{12.44}$$

The Popov line in the $G_p(s)$ plane passes the point $(-1/k_2,\ \ j0)$ with the slope $1/q$.

The condition of absolute stability(12.37) is satisfied if the position of the $G_p(j\omega)$ plot is to the right of the Popov line, i.e., if the Popov line does not intersect the $G_p(j\omega)$ plot in figure12.8. Comparing the Popov characteristics $G_p(j\omega)$ and frequency characteristics $G_L(j\omega)$ of the linear part of the system, the following features can be observed:

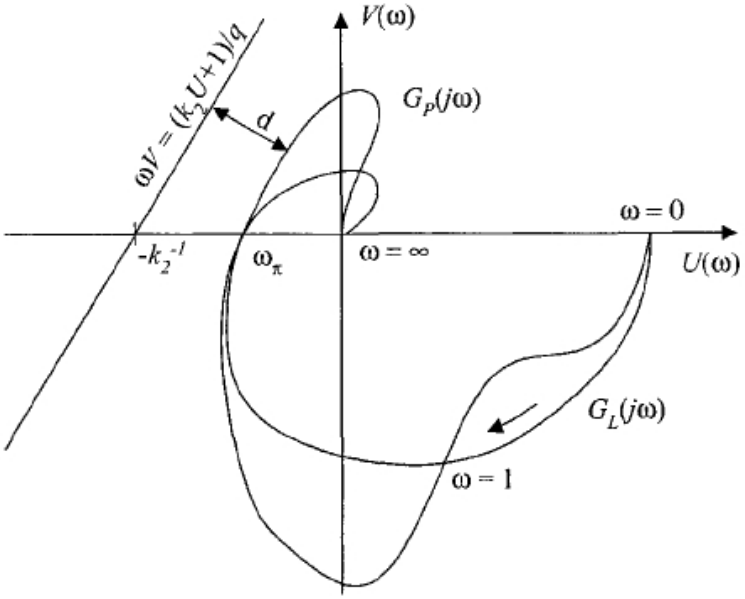


Figure12.8 Popov line does not intersect Popov curve-graphical condition of absolute stability

(1) $G_p(j\omega)$ and $G_L(j\omega)$ intersect the real axis at the same point $\omega = \omega_\pi$.

(2) $\mathrm{Im}\{G_p(j\omega)\} = \omega V(\omega) = V_p(\omega)$ is an even function of frequency ω, while $\mathrm{Im}\{G_L(j\omega)\} = V(\omega)$ is an odd function of frequency ω; the $G_p(j\omega)$ plot is not symmetric with respect to the real axis, while $G_L(j\omega)$ plot is symmetric for $\omega = -\omega$.

(3) $G_p(j\omega)$ plot starts for $\omega = 0$ always from the real axis of the complex plane, while $G_L(j\omega)$ plot can have the starting point on the imaginary axis.

(4) If $\lim_{\omega\to\infty} G_L(j\omega) = 0$, $\lim_{\omega\to\infty} G_p(j\omega)$ can be equal either to zero or to some other final value.

In cases when the Popov plot has a non-convex form, i.e., when it is of much more complex form, the criterion of absolute stability is much more strict. K_{crit} of a convex plot can be much greater than $k_{2\max}$ of a non-convex plot, which could mean that the gain of the nonlinear system with a convex Popov plot can be greater than that of the nonlinear system which has a non-convex Popov plot.

Convex forms for Popov plots represent nonlinear systems where the linear part of the system contains cascaded inertial and oscillatory terms and no more than one integral component, with the condition that the damping ratio of the oscillatory terms is $\xi > \sqrt{2}/2$.

Convex plots $G_p(j\omega)$ also represent nonlinear systems with the following linear parts:

$$G_L(s)=Ke^{-s\gamma};$$
$$G_L(s)=Ks^{-1}e^{-s\gamma};$$
$$G_L(s)=K\prod_{i=1}^{n}(T_is+1)^{-1}; \qquad n\le 6$$
$$G_L(s)=Ks^{-1}\prod_{i=1}^{n}(T_is+1)^{-1}; \qquad n\le 5$$
$$G_L(s)=K\prod_{i=1}^{n}(T_is+1)^{-1}\left(\tau_i s^2+2\xi\tau_i s+1\right)^{-1}; \quad n\le 4; \quad \xi\ge\sqrt{2}/2$$

Instead of the inequality(12.43) which contains the variable quantities q and ω, the Popov criterion can be expressed by one variable quantity only — ω. Then it is more appropriate to determine k_2 analytically:

$$\Theta=\alpha_{\max}-\alpha_{\min}<\pi \tag{12.45}$$

Here

$$\alpha(\omega)=\arg\left[G_p(j\omega)+\frac{1}{k_2}\right]$$

$\alpha_{\max}$ and $\alpha_{\min}$ are maximal and minimal values of $\alpha(\omega)$ in the region $0<\omega<+\infty$. $\alpha(\omega)$ represents the argument of a complex number $\left[G_p(j\omega)+1/k_2\right]$ when $0<\omega<\infty$. For some ω_1, $\alpha(\omega_1)$ will be the angle of a phasor with the real axis, starting at the point $-1/k_2$ and with the peak at $G_p(j\omega_1)$. From equation(12.45) it is obvious that the absolute stability of the nonlinear system can be determined without knowing the exact value of the parameter q — it is enough to draw the Popov line through the point $(-1/k_2,\ j0)$ for at least one slope q with the condition that $0<q<\infty$.

4. Absolute stability with unstable linear part

When the linear part of the system is unstable, it is necessary to accomplish its stabilization with linear feedback in figure12.7. The equivalent transfer function of the linear part of the system $G_E(s)$ is then given by the following

$$G_E(j\omega)=\frac{U(1+K_rU)+K_rV^2}{(1+K_rU)^2+(K_rV)^2}+j\frac{V}{(1+K_rU)^2+(K_rV)^2}=U_E(j\omega)+jV_E(j\omega) \tag{12.46}$$

The equivalent single-valued time-invariant nonlinear function must satisfy the following conditions:

$$F_E(0)=0 \tag{12.47a}$$
$$xF_E(x)>0; \quad x\ne 0 \tag{12.47b}$$
$$\int_0^{\pm\infty}F_E(x)dx=\pm\infty \tag{12.47c}$$
$$0<\frac{F_E(x)}{x}<K_F; \quad x\ne 0 \tag{12.47d}$$

Here, $K_F = k_2 - K_r$ is the value of the new and smaller slope of the sector $[0, \quad K_F]$ inside which the equivalent nonlinear function $F_E(x)$ may be situated. K_r is the stabilizing linear operator in the feedback of the unstable linear part of the system.

For the system with unstable linear part, the absolute stability of the nonlinear system is expressed by the following theorem:

Theorem12.3.2: Popov criterion of absolute stability— G_L stabilized

The equilibrium state of an unforced nonlinear control system with the structure as in figure12.7 will be globally asymptotically stable—absolutely stable if the following is true:

(1) The linear part of the system is time-invariant, stabilized(equivalent linear part of the system is stable) and completely controllable.

(2) Time invariant single-valued nonlinear function $F_E(x)$ is of the class $[0, \quad K_F]$.

(3) There exist two strictly positive numbers $q > 0$ and arbitrarily small number $\delta > 0$ such that for all $\omega \geq 0$ the following inequality is valid:

$$\mathrm{Re}\{(1 + jq\omega)G_E\} + \frac{1}{K_F} \geq \delta > 0 \tag{12.48}$$

Or

$$\mathrm{Re}\{(1 + jq\omega)G_E\} + \frac{1}{K_F} > 0 \tag{12.49}$$

Here, $0 < K_F < \infty; \quad K_F = k_2 - K_r < \infty; \quad \lim_{\omega \to +\infty} G_E(j\omega) = 0$

$$G_E(j\omega) = [G_E(s)]_{s=j\omega} = U_E(\omega) + jV_E(\omega)$$

The graphical interpretation of theorem12.3.2 is the following:

Inserting expression(12.46) into the inequality(12.49) (Popov condition) — and rearranging terms, the conditions for absolute stability are obtained, and graphical interpretation is possible. From equations(12.46) and (12.49), we can get

$$U(1 + K_r U) + K_r V^2 - q\omega V + \frac{(1 + K_r U)^2 + (K_r V)^2}{K_F} \geq 0 \tag{12.50}$$

Or

$$U^2 + \frac{K_F + 2K_r}{K_r(K_r + K_F)}U + V^2 - \frac{qK_F}{K_r(K_r + K_F)}\omega V + \frac{1}{K_r(K_r + K_F)} \geq 0 \tag{12.51}$$

For $\omega > 0$ and $V^2 > 0$, the inequality(12.51) can be quite well approximated by

$$V_p(\omega) < \frac{K_r(K_r + K_F)}{qK_F}U^2(\omega) + \frac{2K_r + K_F}{qK_F}U(\omega) + \frac{1}{qK_F} \tag{12.52}$$

Here, $K_r > 0; \quad K_F = k_2 - K_r; \quad V_p(\omega) = \omega V(\omega) = \mathrm{Im}\{G_p(j\omega)\}$. The inequality(12.52) can be graphically interpreted in the following manner:

In order that the nonlinear system is absolutely stable, the Popov plot must be outside of the parabola with peak at the point S with coordinates:

$$\left[-\frac{(K_F+2K_r)}{2K_r(K_F+K_r)},\quad -\frac{0.25K_F}{qK_r(K_F+K_r)}\right]$$

12.3.2 Nonlinear system controllability

In this section, we will discuss the nonlinear system controllability on the basis of Chapter5. Now let's together consider smooth affine nonlinear control system:

$$\begin{aligned}\dot{x} &= f(x,\ u),\quad x\in R^n,\quad u\in R^m\\ y &= h(x),\quad y\in R^p\end{aligned}\tag{12.53}$$

Here x is the state variable, y is the output, and u is the control.

In equation(12.53), $f(x,\ u)$ is a smooth mapping. In most cases it is assumed that $f(0,\ 0)=0$. Depending on the context"smooth" could mean C^r, C^∞ or C^ω respectively. Mostly, $h(x)$ is also assumed to be smooth(as smooth as $f(x,\ u)$) with $h(0)=0$. u can be assumed to be piecewise continuous or measurable or smooth as we wish. (In fact, it does not affect the controllability and observability much) When $u=u(t)$, it is called an open-loop control; when $u=u(x)(u=u(y))$ it is called a state feedback(output feedback) closed-loop control.

Generally, the state, control and output spaces may be replaced by n, m, *and* p dimensional manifolds respectively.

A particular form of system(12.53) is affine nonlinear system, which is nonlinear in state x and linear in control u . An affine nonlinear control system is generally described as

$$\begin{aligned}\dot{x} &= f(x)+\sum_{i=1}^{m} g_i(x)u_i := f(x)+g(x)u,\quad x\in R^n,\quad u\in R^m\\ y &= h(x),\quad y\in R^p\end{aligned}\tag{12.54}$$

The affine nonlinear systems are the main object in this book. In fact, it is also the main object of nonlinear control theory. In the geometric approach, the $f(x)$ and $g_i(x),\ i=1,\cdots,m$ in system(12.54) are considered as smooth vector fields on R^n .

Now consider system(12.54). The state space, M , is assumed to be R^n or any path connected n dimensional differentiable manifold. Let the control, $u(t)$, be measurable functions. For the ease of statement, we also assume that for any feasible control u the vector field $f(x,\ u)$ is complete. Although this assumption is not necessary for the following discussion on controllability, it serves the purpose of avoiding discussion of the existence of solution over t .

Definition12.3.2.1

Consider system(12.53) and make $x_0\in M$. If the reachable set $R(x_0)=M$, the

system(12.53) is said to be controllable at x_0. If

$$R(x)=M,\quad \forall x\in M$$

the system(12.53) is said to be controllable.

In many cases it is difficult to get the global properties of a nonlinear system, hence we turn to consider the local situation. As for local controllability, let $x_0\in M$ and U be a neighborhood of x_0. A point x_1 is said to be U-reachable, if $x_1\in U$ and there exists a feasible control u and a moment $T>0$, such that the trajectory $x(t)$ satisfies $x(0)=x_0,\quad x(T)=x_1$ and $x(t)\in U,\quad 0\le t\le T$.

The U-reachable set of x_0 is denoted by $R_U(x_0)$.

Definition12.3.2.2

Consider system(12.53) and let $x_0\in M$. If for any neighborhood U of x_0 the reachable set $R_U(x_0)$ is also a neighborhood of x_0, system(12.53) is said to be locally controllable at x_0. If the system is locally controllable at each $x\in M$, system(12.53) is said to be locally controllable.

Neither controllability nor local controllability is symmetric with respect to any two ending points. That is to say, for two given points $x_1,x_2\in M$, $x_2\in R(x_1)$ does not mean $x_1\in R(x_2)$. Similarly, there exists a neighborhood U of x_1 such that $x_2\in R_U(x_1)$ does not mean $x_1\in R_U(x_2)$. We give a symmetric definition as follows:

Definition12.3.2.3

Assume $x_0,x_T\in M$ be given. If there exist $x_1,\cdots,x_k=x_T$, such that or $x_i\in R(x_{i-1})$, or $x_{i-1}\in R(x_i),\quad i=1,\cdots,k$, then x_T is said to be *weakly reachable* form x_0.

The weakly reachable set of x_0 is denoted by $WR(x_0)$. Similarly, we can define a locally weakly reachable set as

Definition12.3.2.4

For system(12.53), x_T is said to be locally weakly reachable from x_0 with respect to a neighborhood, U, if there exist $x_0,x_1,\cdots,x_k=x_T$, such that the trajectories $x_s(t)$ either from x_{s-1} to x_s, or from x_s to x_{s-1}, are contained in U, $s=1,\cdots,k$.

The locally weakly reachable set of x_0 with respect to U is denoted as $WR_U(x_0)$.

Definition13.3.2.5

System(12.53) is weakly controllable at $x_0\in M$ if the weakly reachable set $WR(x_0)=M$. The system is said to be weakly controllable if it is weakly controllable everywhere, i.e., $WR(x)=M,\quad \forall t\in M$. The system is said to be locally weakly controllable at $x_0\in M$, if for any neighborhood U of x_0 the locally weakly reachable set $WR_U(x_0)$ is still a neighborhood of x_0. If system(12.53) is locally weakly controllable

at every $x \in M$, the system is said to be locally weakly controllable.

Proposition12.3.2.1

If system(12.53) is locally weakly controllable, then it is weakly controllable.

Demonstration

For any two points x_0, x_T draw a path $P = \{p(t) | 0 \le t \le 1\}$, connecting x_0 and x_T. Using local weak controllability, for each point $x = p(t)$, $0 \le t \le 1$, there exists an open neighborhood $V_x \subset WR_U(x)$. Then

$$\cup_{x \in P} V_x \supset P$$

It is an open covering of the compact set P. Thus we can find a finite sub-covering $\{V_{x_i} | 1 \le i \le k\}$. By the symmetry of local weak controllability,

$$x_{i+1} \in WR(x_i), \quad i = 0, \cdots, k$$

Here, $x_{k+1} := x_T$. Therefore, $x_T \in WR(x_0)$, which means that the system is weakly controllable.

A point x_T is said to be reachable from x_0 with negative time, denoted as $x_T \in R^-(x_0)$ if there exists a feasible u and $t < 0$ such that $e_t^{f(x,\ u)}(x_0) = x_T$.

Proposition12.3.2.2

Assume the feasible controls are state feedback controls, $u = u(x)$, which are piecewise constant, then

$$R^{\pm}(x_0) = WR(x_0) \tag{12.55}$$

Demonstration

Assume $x_T \in R(x_0)$. That is to say, $x_T = e_t^{f(x,\ u)}(x_0)$. Then $x_0 = e_{-t}^{f(x,\ u)}(x_T)$. Now assume $x_T \in R^{\pm}(x_0)$, namely there exist x_1, $\cdots$, $x_k = x_T$, u_i, δ_i, such that $e_{\delta_i}^{f(x,\ u_i)}(x_i) = x_{i+1}$, $i = 0,1,\cdots,k$. If $\delta_i > 0$, $x_{k+1} \in R(x_k)$; and if $\delta_i < 0$, $x_k \in R(x_{k+1})$. That is, $x_T \in WR(x_0)$. Similarly, $x_T \in WR(x_0)$ implies $x_T \in R^{\pm}(x_0)$.

In the following we assume that the feasible controls are piecewise constant.

Define a set of vector fields as

$$F = \{f(x,\ u) | u = const\}$$

Definition13.3.2.6

The Lie algebra generated by F, $\{F\}_{LA}$ is called the accessibility Lie algebra. Consider $\{F\}_{LA}$ as a distribution. If

$$\dim(\{F\}_{LA})(x_0) = n$$

then it is said that the accessibility rank condition of the system(12.53) is satisfied at x_0. If the accessibility rank condition is satisfied at every $x \in M$, it is said that for system(12.53) the accessibility rank condition is satisfied.

Proposition12.3.2.3

System(12.53) is controllable, if (i) M is simply connected; (ii) all the vector fields in F are complete;(iii) if $X \in F$, then $-X \in F$; and (iv) the accessibility rank condition is satisfied.

Demonstration

Note that condition(iii) implies that

$$R^{\pm}(x) = R(x), \quad \forall x \in M$$

Then the negative t is allowed. Using Chow's theorem, the conclusion follows.

Example12.6 For an affine nonlinear system(12.54), if the drifting term $f(x) = 0$, then condition(iii) in the above proposition12.3.3 is satisfied automatically. Thus if M is simply connected; g_i; $i = 1, \cdots, m$ are complete; and $\dim\{g_1, \cdots, g_m\}_{LA} = n$, the system is controllable.

In general, to verify controllability of a nonlinear control system is hard work. Thus we consider some weaker types of controllability. Weak controllability and local weak controllability are two of them.

We consider the relationship between weak controllability and accessibility rank condition.

Theorem12.3.2.1

Consider system(12.53), if the accessibility rank condition is satisfied at x_0, the system is locally weakly controllable at x_0.

Demonstration

Since $rank(\{F\}_{x_0}) = n$, there exists a neighborhood U of x_0 such that $rank(\{F\}_x) = n, \quad x \in U$. Now if $X(x_0) = 0, \quad \forall X \in F$, it is obvious that $rank(\{F\}_x) = 0$. Therefore, there exists a $f_1 \in F$ such that $f_1(x) \neq 0, \quad x \in U_1 \subset U$. Then we have an integral curve, $e_{t_1}^{f_1}(x_0), \quad -\varepsilon_1 < t_1 < \varepsilon_1$, which is a one dimensional sub-manifold, denoted by L_1. We then claim that there exists a $-\varepsilon_1 < t_1^0 < \varepsilon_1$, and $f_2 \in F$ such that at t_1^0, $f_2 \notin T_{x_1}(L_1)$, where $x_1 = e_{t_1^0}^{f_1}(x_0)$. Otherwise $\dim(\{F\}(x)) = 1 \quad x \in L_1$. Now we consider the mapping

$$\pi_2(t_2, \quad t_1) := e_{t_2}^{f_2} e_{t_1}^{f_1}(x_0)$$

Since the following Jacobian matrix is nonsingular

$$J_{\pi_2}(0, \quad t_1^0) = (f_2(x_1), \quad f_1(x_1))$$

locally this is a diffeomorphism from $|t_2| < \varepsilon_2$, $|t_1 - t_1^0| < e_1$ to the image of π_2. The image is a two dimensional manifold, denoted by L_2. Using the above argument again, we can show that there exists $|t_2^0| < \varepsilon_2$ and $|\bar{t}_1^0 - t_1^0| < e_1$, such that for a $f_3 \in F, \quad f_3(x_2) \notin T_{x_2}(L_2)$,

where $x_2 = e^{f_2}_{t_2^0} e^{f_1}_{t_1^0}(x_0)$. For notational ease, we still use t_1^0 for $\bar{t}_1^0$ and define

$$\pi_3(t_3, \quad t_2, \quad t_1) = e^{f_3}_{t_3} e^{f_2}_{t_2} e^{f_1}_{t_1}(x_0)$$

We obtain a three dimensional sub-manifold L_3 this way. Continuing this procedure, we can finally find $f_1, \cdots, f_n \in F$ and construct a local diffeomorphism from a neighborhood V_n of $(0, \quad t_{n-1}^0, \quad \cdots, \quad t_1^0) \in R^n$ to a neighborhood U_n of $e^{f_n}_0 \cdots e^{f_2}_{t_2^0} e^{f_1}_{t_1^0}(x_0)$.

$$\pi_n(t_n, \quad \cdots, \quad t_2, \quad t_1) = \pi_2(t_2, \quad t_1) := e^{f_n}_{t_n} \cdots e^{f_2}_{t_2} e^{f_1}_{t_1}(x_0) \tag{12.56}$$

In addition, we construct a diffeomorhism which maps U_n back to a neighborhood of x_0. Finally, we define

$$G(x) = e^{f_1}_{-t_1^0} e^{f_2}_{-t_2^0} \cdots e^{f_{n-1}}_{-t_{n-1}^0}(x), \quad x \in U_n$$

Finally, we define the following expression which is a local deiffomorphism from V_n to a neighborhood of x_0.

$$\pi := G \circ \pi_n(t), \quad t \in V_n$$

By definition

$$\pi(V_n) \subset WR_U(x_0)$$

the system is locally weakly controllable at x_0.

Conversely, if the system is locally weakly controllable or even weakly controllable, we would like to know whether that the accessibility rank condition is satisfied.

Proposition12.3.2.4

Assume that system(12.53) is locally weakly controllable, then there exists an open dense set $D \subset M$ such that the accessibility rank condition is satisfied on D, i.e.,

$$\dim\{F\}\big|_x = n, \quad x \in D$$

Proof. If the accessiblity rank condition is satisfied at a point x_0, then there is a neighborhood, U_{x_0} of x_0, such that the accessibility rank condition is satisfied at all $x \in U_{x_0}$. So, it is obvious that the set of points, D, where the accessibility condition is satisfied, is an open set. Now assume D is not dense. Then there is an open set $U \neq \varnothing$, such that

$$\dim\{F\}\big|_x < n, \quad x \in U$$

Let k be the largest dimension of F on U, i.e.,

$$k = \max_{x \in U} \dim\{F\}\big|_x < n$$

Then there exists a non-empty open subset $V \subset U$, such that

$$\dim\{F\}\big|_x = k, \quad x \in V$$

According to Frobenius' theorem, for any $x_0 \in V$, there exists an integral submanifold

of F, $I(F, \quad x_0)$, which has dimension k. It is obvious that

$$WA_V(x_0) \subset I(F, \quad x_0) \cap V$$

which contradicts to the local accessibility of x_0.

For analytic case, accessibility implies the accessibility rank condition.

Proposition12.3.2.5

Assume system(12.53) is analytic and weakly controllable, then the accessibility rank condition is satisfied, i.e.,

$$\dim\{F\}|_x = n, \quad x \in M$$

Demonstration

Assume

$$\dim\{F\}|_x < n, \quad \forall x \in M$$

Then for any $x_0 \in M$ the integral sub-manifold has $\dim(I(F, \quad x_0)) < n$. But

$$WA(x_0) \subset I(F, \quad x_0)$$

It contradicts with weak controllability. Hence there is at least one point $x_0 \in M$, such that

$$\dim\{F\}|_{x_0} = n$$

Next we claim that if the system is weakly controllable, then for any two points $x, y \in M$,

$$\dim\{F\}|_x = \dim\{F\}|_y$$

if we can provide the claim then the proof is done. Using the definition of accessibility, we see that there exist $x_0 = x, x_1, \cdots, x_k = y$, such that x_i and x_{i+1} can be connected by an integral curve of $X \in F$,i.e.,

$$x_{i+1} = e_t^X(x_i)$$

Now for any $Y \in F$, using the Campbell-Baker-Hausdorff formula we have

$$(e_t^X) * Y(x_{i+1}) = \sum_{k=0}^{\infty} ad_X^k Y(x_i) \frac{t^k}{k!} \tag{12.57}$$

Since the right hand side of equation(12.57) is in $F(x_i)$ and the $Y \in F$ is arbitrary

$$(e_{-t}^X) * F|_{x_{i+1}} \subset F|_{x_i}$$

Since $(e_{-t}^X)*$ is a diffeomorphism, it is clear that

$$\dim\{F\}|_{x_{i+1}} \le \dim\{F\}|_{x_i}$$

Exchanging x_i with x_{i+1} and using Campbell-Baker-Hausdorff formula again we can get

$$\dim\{F\}|_{x_i} \le \dim\{F\}|_{x_{i+1}}$$

It follows that

$$\dim\{F\}\big|_x = \dim\{F\}\big|_{x_1} = \cdots = \dim\{F\}\big|_y$$

Then the following corollary is an immediate consequence of Proposition12.3.1, Theorem12.3.1 and Proposition12.3.5.

Corollary12.3.2.1

For an analytic nonlinear system of the form of (12.53), local weak controllability is equivalent to the accessibility rank condition.

Next, let's consider the real reachable set. Denote the reachable set of $x_0 \in M$ at time t by $R(x_0, \ t)$. It is obvious that

$$R(x_0) = \cup_{t\geq 0} R(x_0, \ t)$$

Definition12.3.2.7

(1) System(12.53) is said to be accessible at $x_0 \in M$, if $R(x_0)$ contain a non-empty open set. The system is said to be accessible if it is accessible at every $x \in M$.

(2) System(12.53) is said to be strongly accessible at $x_0 \in M$, if for any $T > 0$, $R(x_0, \ T)$ contains a non-empty open set. The system is said to be strongly accessible if it is strongly accessible at every $x \in M$.

Theorem12.3.2.2

1. For system(12.53) assume the accessibility rank condition is satisfied at $x_0 \in M$, i.e.,

$$\dim\{F\}\big|_{x_0} = n \tag{12.58}$$

Then it is accessible at x_0. Moreover, for any $T > 0$ the reachable set has an empty interior.

$$\cup_{0\leq t\leq T} R(x_0, \ t) \tag{12.59}$$

2. If the system is analytic, then the accessibility rank condition(12.58) is necessary and sufficient for the system to be accessible at x_0.

3. Assume that the system is analytic and accessibility rank condition(12.58) holds. Then the set of interiors is dense in the set $\cup_{0\leq t\leq T} R(x_0, \ t)$.

Proof is neglected here. Readers can consult related controllability books.

12.3.3 Nonlinear system observability

For system(12.53), if in the state space there are two points x_1 and x_2 such that for any feasible control u the outputs are identically the same, i.e.,

$$y(t, \ x_1, \ u) = y(t, \ x_2, \ u), \quad t \geq 0, \quad \forall u$$

Then x_1 and x_2 are said to be in-distinguishable. The set of in-distinguishable points with respect to x_0 is denoted by $ID(x_0)$.

For a linear system(2.150), observability is independent of the control(see equations (5.38) and (5.48)). The system output expression is as follows

$$y = Ce^{A(t-t_0)}x_0 + C\int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau$$

This implies

$$y_2 - y_1 = Ce^{A(t-t_0)}(x_2 - x_1)$$

Above expression is independent to input control u .

Similar to the case of controllability, global observability is also a difficult topic in the investigation of nonlinear systems. For this reason we are primarily interested in local observability.

For a point $x_0 \in M$ and a neighborhood U of x_0, given a $T > 0$, a U feasible control u is such a control that the trajectory remains in U, i.e.,

$$x(t, \quad x_0, \quad u) \in U, \quad \forall t \in [0, \quad T].$$

If for any $T > 0$ and any corresponding U feasible control u,

$$y(t, \quad x_1, \quad u) = y(t, \quad x_2, \quad u), \quad t \in [0, \quad T]$$

then x_1 and x_2 are said to be U -in-distinguishable. The set of U -in-distinguishable points of x_0 is denoted by $ID_U(x_0)$.

Definition12.3.3.1

System(12.53) is said to be locally observable at x_0 if for any neighborhood U of x_0 the U -in-distinguishable set consists of only one point, i.e.,

$$ID_U(x_0) = \{x_0\}, \quad \forall U$$

The system is said to be locally weakly observable at x_0 if there exists a neighborhood U of x_0 the U -in-distinguishable set consists only one point, i.e.,

$$ID_U(x_0) = \{x_0\}$$

It is worth noting that local observability is very strong property. In fact, it implies observability. So we are mostly interested in local weak observability. To investigate the observability we can construct a set of output related functions:

$$H = \left\{ \sum_{i=1}^{s} \lambda_i L_{X_1^i} \cdots L_{X_{k_i}^i}(h_j) \middle| s, k_i < \infty, \quad 1 \le j \le m, \quad \lambda_i \in R, \quad X_k^i \in F \right\} \tag{12.60}$$

Using H we define a co-distribution as

$$H_{co} = \{dh | h \in H\}$$

H_{co} is called the observability co-distribution.

Definition12.3.3.2

System(12.53) is said to satisfy the observability rank condition at x_0 if

$$\dim\{H_{co}\}\big|_{x_0} = n \tag{12.61}$$

If for every $x \in M$, expression(5.33) is satisfied, the system is said to satisfy the observability rank condition.

Next, we consider the relationship between observability and local weak observability.

Theorem12.3.3.1

If system(12.53) satisfies the observability rank condition at x_0, then it is locally weakly observable at x_0.

To prove this theorem, we need the following lemma.

Lemma12.3.3.1

Let $V \subset M$ be an open set. If there exist $x_1, x_2 \in V$ such that

$$x_1 \in ID_V(x_2) \tag{12.62}$$

Then for any function $c(x) \in H$,

$$c(x_1) = c(x_2)$$

Proof.

Choosing any $X_1, \cdots, X_k \in F$, we construct

$$\phi_k(x) = e^{X_1}_{t_1} \cdots e^{X_k}_{t_k}(x)$$

When $\|t\|$ is small enough, we have

$$h_j(\phi_k(x_1)) = h_j(\phi_k(x_2)), \quad j = 1, \cdots, m \tag{12.63}$$

Differentiating both sides with respect to t_1 we have

$$L_{X_1} h_j\left(e^{X_1}_{t_1} \cdots e^{X_k}_{t_k}(x_1)\right) = L_{X_1} h_j\left(e^{X_1}_{t_1} \cdots e^{X_k}_{t_k}(x_2)\right)$$

Setting $t_1 = 0$ yields

$$L_{X_1} h_j\left(e^{X_2}_{t_2} \cdots e^{X_k}_{t_k}(x_1)\right) = L_{X_1} h_j\left(e^{X_2}_{t_2} \cdots e^{X_k}_{t_k}(x_2)\right)$$

Continuing this procedure with respect to t_2, t_3, $\cdots$, t_k, we finally have

$$L_{X_k} L_{X_{k-1}} \cdots L_{X_1} h_j(x_1) = L_{X_k} L_{X_{k-1}} \cdots L_{X_1} h_j(x_2)$$

Now let's together prove theorem12.3.3.1

Proof.

Since $\dim H_{co} = n$, we can choose n functions $c_i(x), \quad i = 1, \cdots, n$, such that they are linearly independent at x_0. Define a mapping

$$P(x) = (c_1(x), \quad c_2(x), \quad \cdots, \quad c_n(x))^T$$

By construction there exists a neighborhood U of x_0 such that $P: U \to P(U) \subset R^n$ is a diffeomorphism. Now for any $x_0 \neq x \in U$, since $P(x) \neq P(x_0), \quad x \notin ID_U(x_0)$.

Hence

$$ID_U(x_0) = x_0$$

Now a natural question is if the observability rank condition is necessary for local weak observability. We have the following results.

Theorem12.3.3.2

If system(12.53) is locally weakly observable, then the observability rank condition is satisfied on an open dense subset of M.

Proof

The set of points where the observability rank condition is satisfied is obviously an open set. So we have only to prove that it is dense.

Assume that there is a non-empty open set U, such that

$$\dim\{H_{co}(x)\} = k < n, \quad x \in U$$

In fact, we can assume $k = \max_{x\in U} \dim(H_{co}(x))$. Then there exists an open subset $V \subset U$, such that $\dim\{H_{co}(x)\} = k, \quad x \in V$.

Then we can find k functions in H such that all the co-vector fields in H_{co} can be expressed as a linear combination of the k forms $dc_i(x_i), \quad i = 1, \cdots, k$, locally on U. Constructing a local coordinate chart as $(U, \quad z)$, where $z = (z^1, \quad z^2)$ and $z^1 = (c_1, \quad \cdots, \quad c_k)^T$. Then the system(12.53) can be expressed as

$$\begin{cases} \overset{\bullet}{z}^1 = f^1(z, \quad u) \\ \overset{\bullet}{z}^2 = f^1(z, \quad u) \\ y_j = h_j(z), \quad j = 1, \cdots, p \end{cases} \tag{12.64}$$

Since $dh_j \in H_{co}$, h_j depends on z^1, i.e.,

$$h_j(z) = h_j(z^1).$$

Next, we claim that f^1 depends only on z^1 too, i.e.,

$$f^1(z, \quad u) = f^1(z^1, \quad u)$$

Otherwise, say for some $1 \le i \le k, \quad k+1 \le j \le n$

$$\frac{\partial f_i^1(z, \quad u)}{\partial z_j^2} \neq 0, \quad x \in U.$$

Then the $j-th$ component of $L_{f(z, \ u)}(c_i(x))$ is

$$dL_{f(z, \ u)}(c_i(x))\big|_j = \frac{\partial f_i^1(z, \quad u)}{\partial z_j^2} \neq 0$$

Hence for some $x \in U$,

$$dL_{f(z, \ u)}(c_i(x)) \notin H_{co}(x)$$

which is a contradiction.

Now system(12.64) can be locally expressed as

$$\begin{cases} \dot{z}^1 = f^1(z^1, u) \\ \dot{z}^2 = f^2(z, u) \\ y_j = h_j(z^1), \quad j = 1, \cdots, p \end{cases} \tag{12.65}$$

Choosing two points $z_1, z_2 \in U$ as

$$z_1 = (z_1^1, z_1^2), \quad z_2 = (z_2^1, z_2^2), \quad z_1^1 = z_2^1, \quad z_1^2 \neq z_2^2$$

Then

$$y(z_1, u(t)) = y(z_2, u(t)), \quad z \in U$$

The system is not locally observable at x_0.

Note that in the above proof, we assume that the feasible controls are piecewise constant. If the feasible controls are state feedback controls, the proof remains true(with some mild modification).

Finally we continue to consider analytic case.

Theorem12.3.3.3

If system(12.53) is analytic and satisfies the controllability rank condition. It is locally weakly observable if and only if the observability rank condition is satisfied..

Proof.

Since the controllability rank condition is satisfied, it is locally weakly controllable. That is, for any two points $x, y \in M$, $x \in WR(y)$, Using theorem12.3.3.1, we only have to prove the necessity. Using theorem12.3.3.2, it suffices to show that H_{co} has constant dimension everywhere. Using the local weak controllability, it is enough to show that $\dim(H_{co})$ is the same for any two points connected by an integral curve of $\{F\}$. Let $x_2 = e_t^X(x_1)$ for some $X \in \{F\}$. For any $\omega \in H_{co}$, we use Campbell-Baker-Hausdorff formula to get

$$(e_t^X)^* \omega(x_2) = \sum_{k=0}^{\infty} L_X^k \omega(x_1) \frac{t^k}{k!}$$

Now the left hand side is in $H_{co}|_{x_1}$ and $(e_t^X)^*$ is an isomorphism. It follows that

$$\dim\{H_{co}\}|_{x_2} \leq \dim\{H_{co}\}|_{x_1}$$

Using $(e_{-t}^X)^*$ in reverse time, we also have

$$\dim\{H_{co}\}|_{x_1} \leq \dim\{H_{co}\}|_{x_2}$$

In the following, we consider an unforced nonlinear of the following form

$$\begin{cases} \dot{x} = f(x) \quad f: R^n \to R^n \\ y = h(x) \end{cases} \tag{12.66a}$$

And then we look for observability conditions in a neighborhood of the origin $x=0$.

Theorem12.3.3.4

The state space realization(12.66a) is locally observable in a neighborhood $U_0 \subset D$ containing the origin, if

$$rank\left(\begin{bmatrix} \nabla h \\ \vdots \\ \nabla L_f^{n-1} h \end{bmatrix}\right) = n, \quad \forall x \in U_0 \tag{12.66b}$$

For linear-invariant systems, condition(12.66b) is equivalent to the observability condition(5.38).

Example12.7 Let

$$\begin{cases} \dot{x} = Ax \\ y = Cx \end{cases}$$

Then $h(x)=Cx$ and $f(x)=Ax$. And we have

$$\begin{aligned} \nabla h(x) &= C \\ \nabla L_f h = \nabla\left(\frac{\partial h}{\partial x}\dot{x}\right) &= \nabla(CAx) = CA \\ &\vdots \\ \nabla L_f^{n-1} h &= CA^{n-1} \end{aligned}$$

Therefore, given system in example12.7 is observable if and only if $S_o = \begin{bmatrix} C & CA & CA^2 & \cdots & CA^{n-1} \end{bmatrix}^T$ is linearly independent or ,equivalently if $rank(S_o)=n$.

Note that the local observability of nonlinear system does not imply global observability in general.

Example12.8 Consider the following nonlinear system

$$\begin{cases} \dot{x}_1 = x_2(1-u) \\ \dot{x}_2 = x_1 \\ y = x_1 \end{cases}$$

The form of this nonlinear system is as follows

$$\begin{cases} \dot{x} = f(x) + g(x)u \\ y = h(x) \end{cases}$$

Here,

$$f(x) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \quad g(x) = \begin{pmatrix} -x_2 \\ 0 \end{pmatrix}, \quad h(x) = x_1$$

If $u=0$, we have

$$rank\left\{\begin{bmatrix}\nabla h & \nabla L_f h\end{bmatrix}^T\right\} = rank\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = 2$$

And thus

$$\begin{cases}\dot{x} = f(x)\\ y = h(x)\end{cases}$$

At the origin, given nonlinear system is observable according to definition12.3.3.4. Now let's consider the same system but assume that $u = 1$. Substituting this input function, we obtain the following dynamical equations:

$$\begin{cases}\dot{x}_1 = 0\\ \dot{x}_2 = x_1\\ y = x_1\end{cases}$$

A glimpse at the new linear time-invariant state space realization shows that the observability has been lost.

12.4 Nonlinear Observer

There are several ways to approach the nonlinear state reconstruction problem, depending on the characteristics of the transfer function(the plant). A complete coverage of the subject is outside the scope of this book and my present ability. In this section, we will discuss two rather different approaches to nonlinear observer design, each applicable to a particular class of systems.

12.4.1 Nonlinear observer with linear error dynamics

Motivated by the work on feedback linearization, it is tempting to approach nonlinear state reconstruction using the following three-step procedures

(1) Find an invertible coordinate transformation that linearizes the state space realization.

(2) Design an observer for the resulting linear system.

(3) Recover the original state using the inverse coordinate transformation defined in (1).

More explicitly, suppose that a system of the form is given

$$\begin{cases}\dot{x} = f(x) + g(x,\ \ u) \quad x \in R^n, \quad u \in R\\ y = h(x) \quad y \in R\end{cases} \tag{12.67}$$

There exist a diffeomorphism $T(\bullet)$ satisfying

$$z = T(x), \quad T(0) = 0, \quad z \in R^n \tag{12.68}$$

And such that, after the coordinate transformation, the new state space realization has the following form

$$\begin{cases} \dot{x} = A_0 z + \gamma(y, \quad u) \\ y = C_0 z \quad y \in R \end{cases} \tag{12.69}$$

where

$$A_0 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad C_0 = [0 \quad 0 \quad \cdots \quad 0 \quad 1], \quad \gamma = \begin{bmatrix} \gamma_1(y, \quad u) \\ \gamma_2(y, \quad u) \\ \vdots \\ \vdots \\ \gamma_n(y, \quad u) \end{bmatrix} \tag{12.70}$$

Then under above conditions, an observer can be constructed according to the following theorem.

Theorem12.4.1.1

If there exist a coordinate transformation mapping system(12.67) into the new form(12.68), then defining

$$\dot{\hat{z}} = A_0 \hat{z} + \gamma(y, \quad u) - K\left(y - \hat{z}_n\right), \quad \hat{z} \in R^n \tag{12.71}$$

$$\hat{x} = T^{-1}(z) \tag{12.72}$$

Such that the eigenvalues of matrix $(A_0 + KC_0)$ are in the left half of the complex plane, then $\hat{x} \to x$ as $t \to \infty$.

Proof.

Let $\tilde{z} = z - \hat{z}$, and $\tilde{x} = x - \hat{x}$. We have

$$\dot{\tilde{z}} = \dot{z} - \dot{\hat{z}} = [A_0 z + \gamma(y, \quad u)] - \left[A_0 \hat{z} + \gamma(y, \quad u) - K\left(y - \hat{z}_n\right)\right]$$

$$= (A_0 + KC_0)\tilde{z}$$

If the eigenvalues of matrix $(A_0 + KC_0)$ have negative real part, then we have that $\tilde{z} \to 0$ as $t \to \infty$.

Using expression(12.72), we obtain

$$\tilde{x} = x - \hat{x} = T^{-1}(z) - T^{-1}\left(z - \tilde{z}\right) \to 0 \quad \text{as} \quad t \to \infty.$$

Hence, theorem12.4.1.1 is found.

Example12.7 Consider the following nonlinear dynamical system

$$\begin{cases} \dot{x}_1 = x_2 + 2x_1^2 \\ \dot{x}_2 = x_1 x_2 + x_1^3 u \\ y = x_1 \end{cases} \tag{12.73}$$

and define the coordinate transformation

$$\begin{cases} z_1 = x_2 - \frac{1}{2}x_1^2 \\ z_2 = x_1 \end{cases}$$

In the new coordinate, system(12.73) takes the following form

$$\begin{cases} \dot{z}_1 = -2y^3 + y^3 u \\ \dot{z}_2 = z_1 + \frac{5}{2}y^2 \\ y = z_2 \end{cases}$$

Referring equation(12.69), we have

$$A_0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad C_0 = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \text{and} \quad \gamma = \begin{bmatrix} -2y^3 + y^3 u \\ \frac{5}{2}y^2 \end{bmatrix}$$

Hence, the observer is as follows

$$\dot{\hat{z}} = A_0 \hat{z} + \gamma(y, u) - K\left(y - \hat{z}_2\right)$$

Namely,

$$\begin{bmatrix} \dot{\hat{z}}_1 \\ \dot{\hat{z}}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \end{bmatrix} + \begin{bmatrix} -2y^3 + y^3 u \\ 5y^2/2 \end{bmatrix} + \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \left(y - \hat{z}_2\right)$$

The error dynamics is as follows

$$\begin{bmatrix} \dot{\tilde{z}}_1 \\ \dot{\tilde{z}}_2 \end{bmatrix} = \begin{bmatrix} 0 & -K_1 \\ 1 & -K_2 \end{bmatrix} \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix}$$

Thus, $\tilde{z} \to 0$ for any $K_1, K_2 > 0$.

It should come as no surprise that, as in the case of feedback linearization, this approach to observer design is based on the cancellation of nonlinearities and therefore assumes “perfect modeling”. In general, perfect modeling is never achieved because system parameters can not be identified with arbitrary precision. Thus, in general, the “expected” cancellations will not take place and the error dynamics will not be linear. The result is that this observer scheme is not robust with respect to parameter uncertainties and that convergence of the observer is not guaranteed in the presence of model uncertainties.

12.4.2 Nonlinear observer with Lipschitz systems

In this section we will discuss nonlinear observer design using a Lyapunov approach. For simplicity, we restrict attention to the case of Lipschitz systems, defined below.

$$\begin{cases} \dot{x} = Ax + f(x, \quad u) \\ y = Cx \end{cases} \tag{12.74}$$

Here $A \in R^{n\times n}$, $C \in R^{1\times n}$, and $f : R^n \times R \to R^n$ is Lipschitz in x on an open set $D \subset R^n$, i.e., function f satisfies the following condition:

$$\left\| f(x_1, \quad u^*) - f(x_2, \quad u^*) \right\| \le \gamma \|x_1 - x_2\|, \quad \forall x \in D \tag{12.75}$$

Now consider the following observer structure

$$\dot{\hat{x}} = A\hat{x} + f\left(\hat{x}, \quad u\right) + L\left(y - C\hat{x}\right) \tag{12.76}$$

Here $L \in R^{n\times 1}$. The following theorem shows that, under these assumption, the estimation error converges to zero as $t \to \infty$.

Theorem12.4.2.1

The system(12.74) and the corresponding observer(12.76), if the Lyapunov equation

$$P(A - LC) + (A - LC)^T P = -Q \tag{12.77}$$

Here $P = P^T > 0$, and $Q = Q^T > 0$, they are satisfied with

$$\gamma < \frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} \tag{12.78}$$

Then the observer error $\tilde{x} = x - \hat{x}$ is asymptotically stable.

Proof

$$\dot{\tilde{x}} = \dot{x} - \dot{\hat{x}} = [Ax + f(x, \quad u)] - \left[A\hat{x} + f\left(\hat{x}, \quad u\right) + L\left(y - C\hat{x}\right) \right]$$

$$= (A - LC)\tilde{x} + f(x, \quad u) - f\left(\hat{x}, \quad u\right)$$

To see that $\dot{\tilde{x}}$ has an asymptotically stable equilibrium point at the origin, consider the Lyapunov function(see Chapter4):

$$V\left(\tilde{x}\right) = \tilde{x}^T P \tilde{x}$$

$$\dot{V}\left(\tilde{x}\right) = \dot{\tilde{x}} P \tilde{x} + \tilde{x} P \dot{\tilde{x}} = -\tilde{x}^T Q \tilde{x} + 2\tilde{x}^T P\left[f\left(\tilde{x} + \hat{x}, \quad u\right) - f\left(\hat{x}, \quad u\right) \right]$$

But,

$$\lambda_{\min}(Q)\|x\|^2 \le \left\| \tilde{x}^T Q \tilde{x} \right\|$$

And

$$\left\| \tilde{x}^T P\left[f\left(\tilde{x}+\hat{x}, \quad u\right) - f\left(\hat{x}, \quad u\right)\right]\right\| \le \left\| 2\tilde{x}^T P \right\| \bullet \left\| f\left(\tilde{x}+\hat{x}, \quad u\right) - f\left(\hat{x}, \quad u\right)\right\| \le 2\gamma\lambda_{\max}(P)\|x\|^2$$

Therefore, $\dot{V}$ is negative define, provided that

$$2\gamma\lambda_{\max}(P)\left\|\tilde{x}\right\|^2 < \lambda_{\min}(Q)\left\|\tilde{x}\right\|^2$$

Or, equivalently

$$\gamma < \frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)}$$

Example12.8 Consider the following nonlinear dynamical system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2^2 \end{bmatrix}$$

Setting

$$L = \begin{bmatrix} 0 & 2 \end{bmatrix}^T$$

Then we have that

$$A - LC = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$$

Solving the Lyapunov equation

$$P(A-LC) + (A-LC)^T P = -Q$$

with $Q = E_{unit}$, we obtain

$$P = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$$

which is positive definite. The eigenvalues of matrix P are $\lambda_{\min}(P) = 0.2929$, and $\lambda_{\max}(P) = 1.7071$. We now consider the function f. Denoting

$$x_1 = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad x_2 = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

We have that

$$\|f(x_1) - f(x_2)\|_2 = \sqrt{\left(\varepsilon_2^2 - \mu_2^2\right)^2} = \left|\varepsilon_2^2 - \mu_2^2\right| = \left|(\varepsilon_2 + \mu_2)(\varepsilon_2 - \mu_2)\right|$$
$$\le 2\left|\varepsilon_2\right|\left\|\varepsilon_2 - \mu_2\right| = 2k\left|\varepsilon_2 - \mu_2\right| \le 2k\|x_1 - x_2\|_2$$

For all x satisfying $|\varepsilon_2| < k$. Thus, $\gamma = 2k$ and f is Lipschitz $\forall x = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 \end{bmatrix}^T : |\varepsilon_2| < k$, and we have

$$\gamma = 2k < \frac{1}{2\lambda_{\max}(P)}, \qquad \text{or} \quad k < \frac{1}{6.8284}$$

The parameter k determines the region of the state space where the observer is guaranteed to work. Of course, this region is a function of the matrix P, and so a function of the observer gain L.

12.5 Nonlinear optimization

The optimality conditions for problems with nonlinear constraints are similar in form to those for problems with linear constraints. However, their derivation is more complicated, even though it is based on related principles. The intuition behind the derivation is the same as optimal problem in linear constraints, but in the case of nonlinear constraints different technical tools are required to give substance to this intuition. In addition, nonlinear constraints can give rise to situations that are impossible in the case of linear constraints.

Since nonlinear optimal control is a very complicated and wide topic, we only give some coarse discussion in this section. The optimality conditions for nonlinearly constrained problems form the basis for algorithms for solving such problems, and so are of great importance. However, not all readers may be interested in studying the derivation of these conditions. For this reason, we firstly state the optimality condition together with some examples. Then we coarsely discuss the use of these optimality condition within optimization algorithms. Only then we present the derivation of the optimality conditions.

12.5.1 Optimality conditions for nonlinear constraints

Here, we present the optimality condition separately for problems with equality and inequality constraints. It is straightforward to combine these results into a more general optimality condition.

The problem with equality constraints is written in the general form as follows:

$$\begin{aligned} &\text{Minimize} && f(x) \\ &\text{Subject to} && g_i(x) = 0, \quad i = 1,2,\cdots,m \end{aligned}$$

The problem with inequality constraints is as follows

$$\begin{aligned} &\text{Minimize} && f(x) \\ &\text{Subject to} && g_i(x) \geq 0, \quad i = 1,2,\cdots,m \end{aligned}$$

Here, we assume that all the functions are twice continuously differentiable.

Some additional assumption must be made to ensure the validity of the optimality conditions. We have chosen to assume that a solution x^* to the optimization problem is a "regular" point. In the case of equality constraints this means that the gradients of the

constraints $\{\nabla g_i(x^*)\}$ are linearly independent. In the case of inequality constraints this means that the gradients of the active constraints at x^*, $\{\nabla g_i(x^*): g_i(x^*)=0\}$, are linearly independent.

Example12.9 (Regularity**).** Consider an equality-constrained problem with two constraints:

$$g_1(x)=x_1^2+x_2^2+x_3^2-3=0$$
$$g_2(x)=2x_1-4x_2+x_3^2+1=0$$

The feasible point is $x^*=(1 \quad 1 \quad 1)^T$. The gradients of the constraints at x^* are

$$\nabla g_1(x^*)=(2 \quad 2 \quad 2)^T$$
$$\nabla g_2(x^*)=(2 \quad -4 \quad 2)^T$$

These two gradients are linearly independent, and so feasible point x^* is a regular point.

Now let's consider an inequality-constrained problem with the single constraint.

$$g_1(x_1, \quad x_2)=\left(\frac{1}{2}x_1^2+\frac{1}{2}x_2^2-1\right)^3\geq 0$$

We get that the feasible point is $x^*=(1 \quad 1)^T$. This constraint is binding at this point, and the gradients of the constraints at x^* are

$$\nabla g_1(x^*)=(0 \quad 0)^T$$

Hence feasible point $x^*=(1 \quad 1)^T$ is not a regular point.

According to the form in equation(7.38), we can write out the optimal condition in the form of Lagrange function:

$$L(x, \quad \lambda)=f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)=f(x)+\lambda^T g(x) \tag{12.79}$$

Here λ is a vector of Lagrange multipliers, and g is the vector of constraint functions $\{g_i\}$. We coarsely discuss these conditions below. They are derived in the section12.5.2.

Theorem12.5.1.1 (Necessary conditions, Equality Constraints)

Assume that x^* is a local minimize solution of function f subject to the constraints $g(x)=0$. Assume that $Z(x^*)$ is a null-space matrix for the Jacobian matrix $\nabla g(x^*)^T$. If x^* is a regular point of the constraints, then there exists a vector of Lagrange multipliers λ^* such that

(1) $\nabla_x L(x^*, \quad \lambda^*)=0$, or equivalently $Z(x^*)^T\nabla f(x^*)=0$, and

(2) $Z(x^*)^T\nabla_{xx}^2 L(x^*, \quad \lambda^*)Z(x^*)$ is positive semidefinite.

Theorem12.5.1.2 (Sufficient conditions, Equality Constraints)

Let x^* be a point satisfying $g(x^*)=0$. Let $Z(x^*)$ be a basis for the null space of $\nabla g(x^*)^T$. Suppose that there exists a vector λ^* such that

(1) $\nabla_x L(x^*, \lambda^*)=0$, and

(2) $Z(x^*)^T \nabla^2_{xx} L(x^*, \lambda^*) Z(x^*)$ is positive semidefinite.

Then x^* is a strict local minimal value of function f in the set $\{x: g(x)=0\}$.

This theorem involves the Jacobian matrix $\nabla g(x^*)^T$, which is the matrix of gradients of the constraint functions. For a linear system of equality constraints $Ax=b$, the Jacobian would be equal to A, and so the condition

$$Z(x^*)^T \nabla f(x^*)=0 \tag{12.80a}$$

Or equivalently

$$\nabla_x L(x^*, \lambda^*)=\nabla f(x^*)-\nabla g(x^*)\lambda^*=0 \tag{12.80b}$$

which is analogous to the following condition for the linear constraint

$$Z^T \nabla f(x^*)=0, \text{ or equivalently } \nabla f(x^*)=A^T \lambda^*$$

The second-order conditions are based on the reduced Hessian

$$Z(x^*)^T \nabla^2_{xx} L(x^*, \lambda^*) Z(x^*)$$

These conditions involve the Hessian of the Lagrange optimal expression $L(x, \lambda)$, while in the case of linear constraints they involve the Hessian of the objective f. The second derivatives of linear constraints are zero, however and so

$$\nabla^2_{xx} L(x^*, \lambda^*)=\nabla^2 f(x^*)$$

in this case. Thus, the second-order condition for linearly constrained problems are a special case of the conditions above.

These optimality conditions are demonstrated in the following example.

Example12.10 (Optimal Conditions, Equality Constraints). Consider the following problem

Minimize $f(x)=x_1^2-x_2^2$

Equality constraints: $x_1^2+2x_2^2=4$

Here we have a single constraint $g(x)=x_1^2+2x_2^2-4=0$. The Lagrange function could be given as follows

$$L(x, \lambda)=x_1^2-x_2^2-\lambda(x_1^2+2x_2^2-4)$$

Therefore, an optimal point must satisfy the following equation group together with the feasible requirement.

$$2x_1-2\lambda x_1=0$$
$$-2x_2-4\lambda x_2=0$$

The first equation has two possible solutions: $x_1=0$ and $\lambda=1$. If $x_1=0$, then from

feasible equation we can get $x_2 = \pm\sqrt{2}$. In another case, the second equation implies that $\lambda = -\frac{1}{2}$. On the other hand, if $\lambda = 1$, then from the second equation we get $x_2 = 0$, and from feasible equation we can get $x_1 = \pm 2$. There are four possible solutions:

$$x = \left(0, \quad \sqrt{2}\right)^T, \quad \lambda = -\frac{1}{2};$$
$$x = \left(0, \quad -\sqrt{2}\right)^T, \quad \lambda = -\frac{1}{2};$$
$$x = \left(2, \quad 0\right)^T, \quad \lambda = 1;$$
$$x = \left(-2, \quad 0\right)^T, \quad \lambda = 1.$$

These are all stationary points of $f(x_1, \quad x_2)$. We can determine which are minimal value by examining the Hessian matrix

$$\nabla^2_{xx} L(x, \quad \lambda) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} - \lambda \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 2(1-\lambda) & 0 \\ 0 & -2(1+2\lambda) \end{bmatrix}$$

Now consider the solution $x = \left(0, \quad \sqrt{2}\right)^T$ with Lagrange multiplier $\lambda = -1/2$. Since $\nabla g(x) = \begin{pmatrix} 2x_1 & 4x_2 \end{pmatrix}^T = \left(0 \quad 4\sqrt{2}\right)^T$, we can choose the null-space matrix $Z = Z(x) = \begin{pmatrix} 1 & 0 \end{pmatrix}^T$. Taking $\lambda = -1/2$, we obtain

$$Z^T \nabla^2_{xx} L(x, \quad \lambda) Z = 3 > 0,$$

Hence the reduced Hessian is positive definite and the point is a strict local minimal value of function f. Similarly, the solution $x = \left(0, \quad -\sqrt{2}\right)^T$ is also a strict local minimal value.

If we take the solution $x = \begin{pmatrix} 2, & 0 \end{pmatrix}^T$, $\lambda = 1$, then $\nabla g(x) = \begin{pmatrix} 2x_1 & 4x_2 \end{pmatrix}^T = \begin{pmatrix} 4, & 0 \end{pmatrix}^T$, and we can choose the null-space matrix $Z = \begin{pmatrix} 0, & 1 \end{pmatrix}^T$. The reduced Hessian is $Z^T \nabla^2_{xx} L(x, \quad \lambda) Z = -6 < 0$, and hence the point is a local maximizer of f. A similar conclusion holds for the point $x = \begin{pmatrix} -1, & 0 \end{pmatrix}^T$. For this problem, all feasible points are regular points.

Next, we give out the necessary conditions for problems with inequality constraints. These conditions are sometimes called the Karush-Kuhn-Tuchker conditions, the KKT conditions.

Theorem12.5.1.3 (Necessary conditions, inequality constraints)

Assume that solution x^* is a local minimum point of function $f(x)$ subject to the constraints $g(x) \geq 0$. Let the columns of $Z(x^*)$ form a basis for the null-space of the Jacobian of the active constraints at x^*. If x^* is a regular point for the constraints, then there exists a vector of Lagrange multipliers λ^* such that

(1) $\nabla_x L(x^*, \quad \lambda^*) = 0$, or equivalently $Z(x^*)^T \nabla f(x^*) = 0$,

(2) $\lambda^* \geq 0$,

(3) $\lambda^{*T} g(x^*) = 0$, and

(4) $Z(x^*)^T \nabla^2_{xx} L(x^*, \lambda^*) Z(x^*)$ is positive semidefinite.

The condition $\lambda^{*T} g(x^*) = 0$ is the complementary slackness condition. Since the vectors λ^* and $g(x^*)$ are both non-negative, it implies that $\lambda_i^* g_i(x^*) = 0$ for each i. This means that either a constraint is active, or its associated Lagrange multipliers corresponding to the active constraints are all positive, then we have strict complementary; otherwise, if a Lagrange multiplier corresponding to an active constraint is zero, the constraint is said to be degenerate.

The second-order sufficiency conditions for a local minimum point are stated below.

Theorem12.5.1.4 (Sufficiency conditions, inequality constraints)

Let x^* be a point satisfying $g(x^*) \geq 0$. Suppose that there exists a vector λ^* such that

(1) $\nabla_x L(x^*, \lambda^*) = 0$,

(2) $\lambda^* \geq 0$,

(3) $\lambda^{*T} g(x^*) = 0$, and

(4) $Z_+(x^*)^T \nabla^2_{xx} L(x^*, \lambda^*) Z_+(x^*)$ is positive definite.

Here, Z_+ is a basis for the null-space of the Jacobian matrix of the non-degenerate constraints (the active constraints with positive Lagrange multipliers) at x^* is a strict local minimizer of function f in the set $\{x : g(x) \geq 0\}$.

These optimal conditions are demonstrated by the following example.

Example12.11(Optimal conditions, Inequality constraints). Consider the problem

Minimize $f(x) = x_1$

Inequality constraint $(x_1 + 1)^2 + x_2^2 \geq 1, \quad x_1^2 + x_2^2 \leq 2$

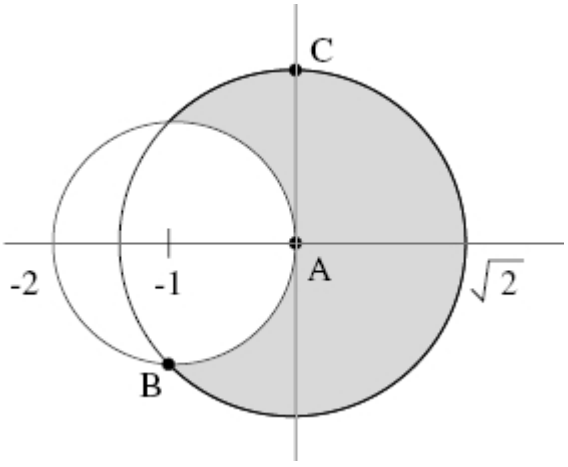


Figure12.9 Problem with nonlinear inequalities

Please test whether the points $A=(0,\quad 0)^T$, $B=(-1,\quad -1)^T$, $C=\left(0,\quad \sqrt{2}\right)^T$ are optimal(see figure12.9).

Rearranging the constraints to the "$\geq$" form we obtain

$$L(x,\quad \lambda)=x_1-\lambda_1\left[(x_1+1)^2+x_2^2-1\right]+\lambda_2\left(x_1^2+x_2^2-2\right)$$

Therefore

$$\nabla_x L(x,\quad \lambda)=\begin{bmatrix}1-2\lambda_1(x_1+1)+2\lambda_2 x_1\\ -2\lambda_1 x_2+2\lambda_2 x_2\end{bmatrix}$$

And

$$\nabla_{xx}^2 L(x,\quad \lambda)=\begin{bmatrix}2\lambda_2-2\lambda_1 & 0\\ 0 & 2\lambda_2-2\lambda_1\end{bmatrix}$$

At the point A, only the first constraint is active, and hence $\lambda_2=0$. Solving for λ_1 we obtain

$$\begin{cases}1-2\lambda_1=0\\ 0=0\end{cases}$$

Then, we get $\lambda_1=\dfrac{1}{2}$.

Therefore, this is a candidate for a local minimizer. Taking $Z=[0\quad 1]^T$ as a basis matrix for the null space of the Jacobian matrix $[2\quad 0]$, we get

$$Z^T\nabla_{xx}^2 L(x,\quad \lambda)Z=(0\quad 1)\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix}=-1$$

And hence the reduced Hessian matrix is negative definite, and the sufficiency conditions are not satisfied. This point is not a local maximizer since $\lambda_1>0$.

At the point B both constraints are active. Solving for the Lagrange multipliers we obtain

$$\begin{aligned}&1-2\lambda_2=0\\ &2\lambda_1-2\lambda_2=0\end{aligned}\quad\Rightarrow\quad \lambda_1=\lambda_2=\frac{1}{2}$$

Therefore the point satisfies the first-order necessary condition for optimality.Moving to the sufficiency conditions, we note that the null-space matrix for the Jacobian is empty.

Therefore the sufficiency conditions are trivially satisfied, and the point is strict local minimal value.

At the point $C=\left(0,\quad \sqrt{2}\right)^T$, only the second constraint is active, and hence $\lambda_1=0$. Solving for λ_2 we obtain

$$1+2\lambda_2(0)=0$$
$$2\lambda_2\sqrt{2}=0$$

This system is inconsistent. Hence the first-order necessary condition is not satisfied and this point is not optimal.

As in the linearly constrained case, the Lagrange multipliers provide a measure of the sensitivity of the optimal objective value to changes in the constraints. This shows up in the optimal conditions which include the requirement that $\lambda^* \geq 0$ for the inequality-constrained problem. The magnitude of the multipliers also has meaning, with a large multiplier indicating a constraint more sensitive to changes in its right-hand side.

The following example shows that if the regularity condition is not satisfied at a local minimizer, the first-order necessary condition for optimality may not hold.

Example12.12(Regularity condition Not satisfied). Consider the problem

Minimize $f(x) = 3x_1 + 4x_2$

Equality constraint $(x_1+1)^2 + x_2^2 = 1, \quad (x_1-1)^2 + x_2^2 = 1$

The solution to this problem is $x^* = (0, \quad 0)^T$, which is also the only feasible point. The gradients of the constraints at x^* are $(2, \quad 0)^T$ and $(-2, \quad 0)^T$, and thus are linearly dependent. Setting the gradient of the Lagrange function with respect to x equal to zero yields

$$3 - 2\lambda_1 + 2\lambda_2 = 0$$
$$4 = 0$$

This is an inconsistent system. Hence there are no multipliers λ_1 and λ_2 for which the gradient of the Lagrange function is zero, even though the point is optimal.

12.5.2 Derivation of Optimal conditions for nonlinear constraints

We now examine the derivation of the optimality conditions for nonlinear constraint problems. We again begin with a problem that has equality constraints only:

Minimize $f(x)$

Subject to $g_i(x) = 0, \quad i = 1,2,\cdots,m$

Each of the functions $f(x)$ and $g_i(x)$ is assumed to be twice continuously differentiable. If we define $g(x)$ as the vector of constraint functions $\{g_i(x)\}$, then the problem is to minimize $f(x)$ limited to $g(x) = 0$. The set of points x such that $g(x)$ is called a surface.

We derive first-order and second-order optimality conditions for this problem. The main difficulty is the characterization of small movements that maintain feasibility. In nonlinear case, this is not possible. For example, consider the nonlinear equality constraint $x_1^2 + x_2^2 = 2$, and let x be any feasible point such as $x = (1, \quad 1)^T$. Any small step taken from x along any direction will result in the loss of feasibility(see figure12.10). Thus there are no feasible directions at this point, or at any other feasible point. To define small

movements that maintain feasibility, we will use feasible curves.

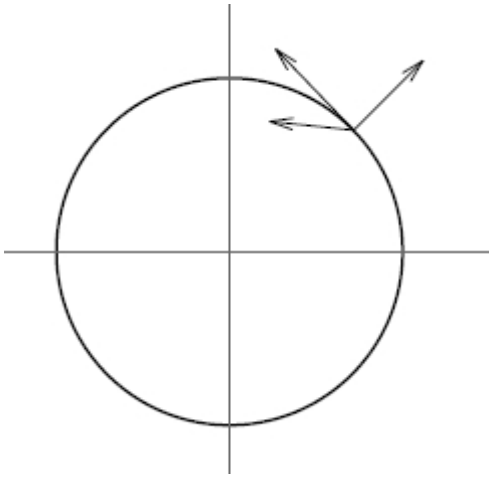


Figure12.10 No feasible direction

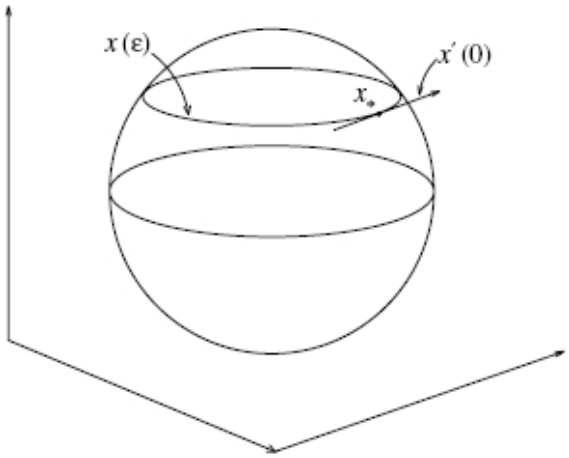


Figure12.11 Feasible curve

Example12.13(Feasible curve). Consider the the following constraint and curve

Constraint $g(x)=x_1^2+x_2^2+x_3^2-3=0$

Curve $x(t)=\begin{bmatrix}\sqrt{2}\cos(t+\pi/4)\\ \sqrt{2}\sin(t+\pi/4)\\ 1\end{bmatrix},\quad -\pi\le t\le\pi$

Then $x(0)=(1,\ 1,\ 1)^T$ and $g(x)=2\cos^2(t+\pi/4)+2\sin^2(t+\pi/4)+1-3=0$. Hence $x(t)$ is a feasible curve passing through the point $(1,\ 1,\ 1)^T$. The tangent to the curve at point $(1,\ 1,\ 1)^T$ is seen in figure12.11.

$$x'(0)=\left.\frac{dx(t)}{dt}\right|_{t=0}=\begin{pmatrix}-\sqrt{2}\sin(t+\pi/4)\\ \sqrt{2}\cos(t+\pi/4)\\ 0\end{pmatrix}=\begin{pmatrix}-1\\ 1\\ 0\end{pmatrix}$$

Now we suppose that solution x^* is a local solution of the optimization problem. Then x^* is a local minimizer of function $f(x)$ along any feasible curve passing through point x^*. Let $x(t)$ be any such curve with $x(0)=x^*$. Then $t=0$ is a local minimizer of the one-dimensional function $f(x(t))$, and the derivative of function $f(x(t))$ with respect to t must vanish at $t=0$. Then we can obtain the following derivative

$$\left.\frac{df(x(t))}{dt}\right|_{t=0}=x'(t)^T\nabla f(x(t))\Big|_{t=0}=x'(0)^T\nabla f(x^*)=0$$

Thus, if point x^* is a local minimizer of function $f(x)$, then

$$x'(0)^T\nabla f(x^*)=0 \quad \text{for all feasible curves } x(t) \text{ through point } x^*. \tag{12.81}$$

Define

$$T(x^*)=\{p: p=x'(0) \textit{ for some feasible curves } x(t) \textit{ through } x^*\}$$

This is the set of all tangents to feasible curves through x^*. Now assume for

convenience that $0 \in T(x^*)$. The set has the property that if $p \in T(x^*)$, then $\alpha p \in T(x^*)$ for any non-negative scalar coefficient α. A set with this property is called a cone, and for this reason $T(x^*)$ is sometimes called the tangent cone at point x^*. The tangent cone at the point $(1, 1, 1)^T$ in example12.13 is shown in figure12.12. It is parallel to the tangent plant at $(1, 1, 1)^T$ but passes through the origin point.

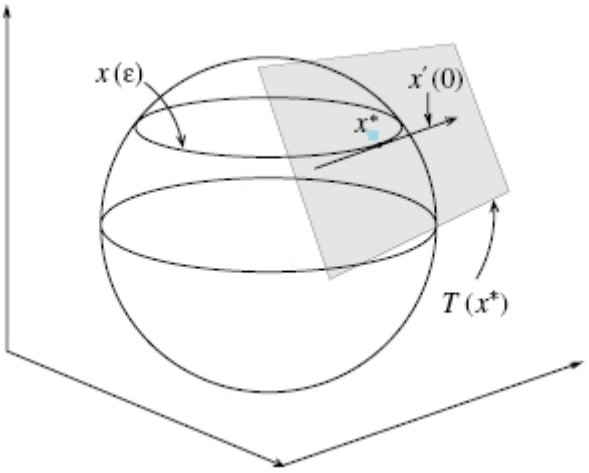


Figure12.12 Tangent cone

From equation(12.81) we obtain a condition for optimality of a feasible point x^*:

$$p^T \nabla f(x^*) = 0 \quad \text{for all} \quad p \in T(x^*)$$

In this form, the optimality condition is not yet practical, since it is not always easy to represent the set of all feasible curves explicitly. We shall develop an alternative characterization of the tangent cone. To this end, we notice that $g_i(x(t))$ is a constant function of time t (it is zero for all t), and hence its derivative with respect to t vanishes everywhere, i.e.,

$$\frac{dg_i(x(t))}{dt} = 0$$

Then we obtain

$$x'(t)^T \nabla g_i(x(t)) = 0$$

In particular, at $t = 0$ we obtain $x'(0) \nabla g_i(x^*) = 0$. Since this is true for all feasible arcs through x^*, we obtain

$$p^T \nabla g_i(x^*) = 0 \quad \text{for all} \quad p \in T(x^*)$$

The equation above holds for each constraint $g_i(x) = 0$. It will be useful to define $A(x^*)$ as the $m \times n$ matrix whose ith row is $\nabla g_i(x^*)^T$. This is the Jacobian matrix of $g(x)$ at point x^*. The equation above can be written as $A(x^*)p = 0$, so that any vector in the tangent cone at x^* also lies in the null space of the Jacobian matrix at x^*:

$$p \in T(x^*) \Rightarrow p \in N(A(x^*))$$

Hence the tangent cone at a point is contained in the null space of the Jacobian matrix

at the point.

Example12.14 (Null space of the Jacobian). Consider the problem in example12.13. At $x^* = (1, \ 1, \ 1)^T$ we have $\nabla g(x^*)^T = (2, \ 2, \ 2)^T$. Thus any vector p in the tangent cone must satisfy $p^T \nabla g(x^*) = 2p_1 + 2p_2 + 2p_3 = 0$, that is to say, $p_1 + p_2 + p_3 = 0$. In this example, the tangent cone and null space of the Jacobian are both equal to the set $\{p : p_1 + p_2 + p_3 = 0\}$.

In the previous example, the tangent cone and the null space of the Jacobian were equal. It can be difficult to characterize the tangent cone, but it is easy to compute the Jacobian matrix and to generate its associated null space. Hence it would be useful if these two sets were always equal. Unfortunately, this is not always the case, as the next example shows.

Example12.15 ($T(x^*) \neq N(A(x^*))$). Consider the the constraint $g(x) = \left(\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 1\right)^2 = 0$. The feasible set is a circle of radius $\sqrt{2}$. The tangent cone at the point $x^* = (1, \ 1)^T$ is the set $T(x^*) = \{p : p_1 + p_2 = 0\}$. Since $\nabla g(x) = \left(2\left(\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 1\right)x_1, \ 2\left(\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 1\right)x_2\right)^T$, the Jacobian matrix at x^* is $A(x^*) = (0, \ 0)$. Therefore the null space of the Jacobian is $N(A(x^*)) = R^2$ and $T(x^*) \neq N(A(x^*))$.

Luckily, example such as above are uncommon. In the majority of problems the tangent cone at a feasible point is indeed equal to the null space of the Jacobian matrix at the point. One condition that guarantees that this is regularity, that is, the assumption that the gradient vectors $\nabla g_i(x^*), \ i = 1,2,\cdots,m$ are linearly independent(or equivalently, that their Jacobian matrix has full row rank). In the next lemma we prove that if point x^* is a regular point, then $T(x^*) = N(A(x^*))$.

Lemma 12.5.2.1

If point x^* is a regular point of the constraints, then $T(x^*) = N(A(x^*))$.

Demonstration

We need only show that $p \in N(A(x^*))$ implies that $p \in T(x^*)$; that is, there exists some feasible curve $x(t)$ through point x^* satisfying $x'(0) = p$.

To prove the existence of a feasible curve we shall use the derivative rule of the implicit function. Let y be an m-dimensional vector, and consider the following system of nonlinear equations in y and t:

$$g(x^* + tp + \nabla g(x^*)y) = 0$$

The system has a solution at $(y, \ t) = (0, \ 0)$. Its Jacobian with respect to y at this

point is

$$\nabla g\left(x^*\right)^T \nabla g\left(x^* + tp + \nabla g\left(x^*\right)y\right)\Big|_{(y,\ t)=(0,\ 0)} = \nabla g\left(x^*\right)^T \nabla g\left(x^*\right)$$

which by the regularity assumption is non-singular. Therefore, by the derivative rule of the implicit function in advanced mathematics, there exists a continuously differentiable function $y = y(t)$ in a neighborhood of $t = 0$ satisfying

$$g\left(x^* + tp + \nabla g\left(x^*\right)y(t)\right) = 0$$

Letting $x(t) = x^* + tp + \nabla g\left(x^*\right)y(t)$, we obtain that $x(t)$ is a feasible curve through x^* with $x(0) = x^*$. It remains only to show that $x'(0) = p$. From the formula for $x(t)$ we have that $x'(0) = p + \nabla g\left(x^*\right)y'(0)$, so we need to show that the second term is zero. Since $x(t)$ is a feasible curve it satisfies $\nabla g\left(x^*\right)^T x'(0) = 0$.Hence

$$\nabla g\left(x^*\right)^T p + \nabla g\left(x^*\right)^T \nabla g\left(x^*\right)y'(0) = 0$$

The first term above is zero because $p \in N\left(A\left(x^*\right)\right)$. The lemma now follows because of the regularity assumption.

If we assume that a local minimizer is a regular point, we can obtain a more useful optimality condition. Let x^* be a local solution that satisfies the regularity condition. Then any vector $p \in N\left(A\left(x^*\right)\right)$ is also in $T\left(x^*\right)$. It follows that

$$p^T \nabla f\left(x^*\right) = 0 \quad \text{for all} \quad p \in N\left(A\left(x^*\right)\right).$$

If $Z\left(x^*\right)$ is a null-space matrix for $A\left(x^*\right)$, then

$$Z\left(x^*\right)^T \nabla f\left(x^*\right) = 0$$

This is the first-order necessary condition for optimality. It states that the reduced gradient at a local minimum must be zero. We note here that the same condition is also satisfied at a local maximum point. The reduced gradient may be also zero at a point that is neither a local maximum nor a local minimum point, that is, at a saddle point.

As in the linear case, we can show that the reduced gradient is zero if and only if there exists an m-dimensional vector λ^* such that

$$\nabla f\left(x^*\right) = A\left(x^*\right)^T \lambda^* = \sum_{i=1}^{m} \lambda_i^m \nabla g_i\left(x^*\right)$$

This is an equivalent statement of the first-order necessary condition for optimality. The coefficients $\left\{\lambda_i^*\right\}$ are the Lagrange multipliers.

We now derive the second-order conditions for optimality. Recall that if point x^* is a local minimizer, then point x^* is a local minimizer along any feasible curve passing through point x^*. Let $x(t)$ be any such curve with $x(0) = x^*$. Then since $t = 0$ is a local minimizer of the function $f(x(t))$, the second derivative of $f(x(t))$ with respect to t must be non-negative at $t = 0$. Using the chain rule, we can obtain

$$\frac{d^2\left(f\left(x(t)\right)\right)}{dt}=\frac{d}{dt}\left[x'(t)^T\nabla f\left(x(t)\right)\right]=x'(t)^T\nabla^2 f(x)x'(t)+\nabla f(x)^T x''(t)$$

Hence

$$\frac{d^2}{dt^2}f\left(x(0)\right)=p^T\nabla^2 f\left(x^*\right)p+\nabla f\left(x^*\right)x''(0)\geq 0$$

where $p=x'(0)$ is the tangent to the curve at x^*. In the expression above, the term $\nabla f\left(x^*\right)^T x''(0)$ does not necessarily vanish. Therefore the second derivative along an arc depends not only on the Hessian of the objective, but also on the curvature of the constraints(that's, on the term $x''(0)$).

To transform this into a more useful condition, it will be convenient to get ride of the term involving $x''(0)$. To do this, we notice that $g_i(x)$ is constant, so its second derivative with respect to time t must vanish for all t, in particular at $t=0$. Using the chain rule we obtain

$$p^T\nabla^2 g_i\left(x^*\right)p+\nabla g_i\left(x^*\right)^T x''(0)=0$$

We can multiply the last equality by λ_i^* and sum over all i. If we subtract the result from the previous inequality, then, because $\nabla_x L\left(x^*,\ \lambda^*\right)=0$, the term involving $x''(0)$ will be eliminated. The final result is that

$$p^T\left[\nabla^2 f\left(x^*\right)-\sum_{i=1}^{m}\lambda_i^*\nabla^2 g_i\left(x^*\right)\right]p\geq 0$$

for all tangent vectors $p\in T\left(x^*\right)$. The term in brackets is the Hessian of L with respect to x at the point $(x,\ \lambda)$. Therefore

$$p^T\left[\nabla_{xx}^2 L\left(x^*,\ \lambda^*\right)\right]p\geq 0$$

for all tangent vectors $p\in T\left(x^*\right)$. Under the regularity assumption, this inequality will hold for any p in $N\left(A\left(x^*\right)\right)$. Consequently, the reduced Hessian $Z\left(x^*\right)^T\nabla_{xx}^2 L\left(x^*,\ \lambda^*\right)Z\left(x^*\right)$ must be positive semidefinite. This the second-order necessary condition for optimality.

The proof of the sufficiency conditions uses similar techniques.

Finally, we consider a problem with nonlinear inequality constraints:

Minimize $f(x)$

Constraint $g_i(x)\geq 0,\quad i=1,2,\cdots,m$

Optimality conditions can be derived by combining the ideas developed for problems with nonlinear equalities with those for problems with linear inequalities. There are a few issues which are unique to problems with nonlinear inequalities. We discuss them briefly.

Let x^* be a feasible solution to the inequality-constrained problem. Whereas in the case of equality constraints we can maintain feasibility by moving in either direction along a feasible curve through point x^*, here it is often possible to move in only one direction;

we shall call this "movement along a feasible arc". More formally, we define an arc emanating from x^* as a directed curve $x(t)$ parameterized by the variable t in an time interval $[0, \quad T]$ for which $x(0)=x^*$. An arc is feasible if $g(x(t))\geq 0$ for t in $[0, \quad T]$. Some examples of feasible arcs are illustrated in figure12.13. The optimality conditions are a result of the requirement that if a small movement is made along a feasible arc, the objective value will not decrease.

The constraints that are inactive at x^* can be ignored, since they do not influence the local optimality conditions. With the regularity assumption, it is possible to derive the first-order and second-order conditions for optimality.

Chapter summary

Nonlinear control is a very complicate topics at present. According to present development status, some topics are coarsely provided and explained such as linearization of nonlinear control system, stability, controllability and observability. Then nonlinear observer and nonlinear optimization are tried to be explained and demonstrated. In this chapter, readers should know the basic concept of nonlinear system, and understand the stability and controllability and observability of nonlinear system. Besides these topics, readers should know how to realize a nonlinear observers. Of course, given topics are coarsely illustrated and shallowly explored for nonlinear system is very complicated and wide topics such as global observability, local control and period controllability, and so on.

Related Readings

[1]Cheng Daizhan, Hu Xiaoming, Shen Tielong.*"Analysis and Design of Nonlinear Control Systems"*,Beijing: Science Press, 2010, China.

[2]H.Nijmeijer, A.J.van der Schaft. *"Nonlinear Dynamical Control Systems"*, New York: Springer, 1990, USA.

[3]Meral Altinay. *"Applications of Nonlinear Control"*, Rijeka: InTech, 2012, Croatia.

[4]Zoran VukiC, Ljubomir KuljaCa, Dali DonlagiC, Sejid TeSnjak. "Nonlinear Control Systems", New York: Marcel Dekker, 2003, USA.

[5]Li Chunbiao, Julien Clinton Sprott, Wesley Thio. *"Linearization of the Lorenz system"*, Physics Letters A, Volume 379, Issues 10–11, 8 May 2015, Pages 888-893.

[6]Hossein Mirzaeinejad, Mehdi Mirzaei. *"Optimization of nonlinear control strategy for anti-lock braking system with improvement of vehicle directional stability on split-μ roads"*, Transportation Research Part C: Emerging Technologies, Volume 46, September 2014, Pages 1-15.

[7]Arnab Maity, Jagannath Rajasekaran, Radhakant Padhi."*Nonlinear control of an air-breathing engine including its validation with vehicle guidance"*, Aerospace Science

and Technology, Volume 45, September 2015, Pages 242-253.
[8]Mirosław Galicki.*"An adaptive non-linear constraint control of mobile manipulators"*, Mechanism and Machine Theory, Volume 88, June 2015, Pages 63-85.
[9]G.A. Leonov, N.V. Kuznetsov, M.V. Yuldashev, R.V. Yuldashev. *"Nonlinear dynamical model of Costas loop and an approach to the analysis of its stability in the large"*, Signal Processing, Volume 108, March 2015, Pages 124-135.
[10]R.M. Brisilla, V. Sankaranarayanan. *"Nonlinear control of mobile inverted pendulum"*, Robotics and Autonomous Systems, Volume 70, August 2015, Pages 145-155.
[11]S. Elloumi, N. Benhadj Braiek. *"On Feedback Control Techniques of Nonlinear Analytic Systems"*, Journal of Applied Research and Technology, Volume 12, Issue 3, June 2014, Pages 500-513.

Review Questions

12.1. What is the basic problem of nonlinear control system?

12.2. What are the basic properties of nonlinear control you know?.

12.3. What methods can be employed to realize the linearization of nonlinear control system?

12.4. Is Lyapounov's stability judgment criterion utilized to judge the stability of nonlinear control system? How is the judgment criterion applied into nonlinear control system?

12.5. Local stability of nonlinear system is the stability of system, is it right?

12.6. Please the main methods that utilize to judge the observability of nonlinear system in terms of your recognition.

12.7. What is the optimality condition for nonlinear constraint system?

12.8. Please give out the connection and difference between nonlinear observer and linear observer.

Problems

Problem12.1. Please find the equilibrium points for the system described by the following differential equation:

$$\ddot{y}+2(1+y)\dot{y}-3y+y^2=0$$

Then evaluate the linearized Jacobian matrix at each equilibrium point and determine the stability characteristics from the eigenvalues.

Problem12.2. The desired response for the second-order nonlinear system described by

$$\dot{x}_1=2x_2^2+u \quad \text{and} \quad \dot{x}_2=x_1u+x_1^2$$

They are intended to imitate the uncoupled linear system